

Université Paris-Est

Graphical Model Inference & Learning for Visual Computing

Nikos Komodakis

Ecole des Pont ParisTech

2013

Mémoire d'habilitation à diriger des recherches

Spécialité: traitement du signal et des images

Membres du jury:

M. Patrick **Bouthemy** (INRIA)

M. Fredrik **Kahl** (Lund University)

M. Patrick **Perez** (Technicolor)

M. Jean-Christophe **Pesquet** (University of Marne-la-Vallée)

M. Christoph **Schnoerr** (Heidelberg University)

M. John **Tsitsiklis** (MIT)

M. Ramin **Zabih** (Cornell University)

Contents

1	Introduction & context	3
2	Scientific foundations	6
2.1	Efficient inference on graphical models based on the duality theory of linear programming	8
2.2	MRF energy minimization and beyond via dual decomposition . .	10
2.3	Tighter LP relaxations and cycle repairing	11
2.4	Accelerating MRF inference	14
2.5	Towards efficient inference for high-order graphical models	18
2.6	Learning of high-order graphical models	21
2.7	Learning of graphical models with weak supervision	26
2.8	Inference and learning for LP-based clustering	27
3	Applications to visual computing	32
3.1	Computer vision and image processing	32
-	Blind image deconvolution using MRF-based image priors . .	32
-	Motion analysis and optical flow estimation	34
-	Knowledge-based image segmentation	35
-	From images to 3D models: visual 3D reconstruction of large natural environments	36
-	Expression mimicking: from video to realistic 3D facial animations	37
-	Tracking	38
3.2	Medical image analysis	38
-	Deformable registration	39
-	Segmentation and reconstruction of anatomical structures . . .	41
-	Groupwise registration	41
-	Learning deformation priors for dense image registration . . .	42
-	Image fusion through deformable mosaicing	42
-	Processing of diffusion tensor images (DTI)	43
4	Acknowledgements	45
5	References	45
	Appendix	57

1 Introduction & context

Visual data are the most complex and most useful sensory input for humans. This is evidenced from the fact that a significant proportion of the human brain is dedicated to vision and to vision only. Thanks to visual perception we are able to sense the light emanating from the objects surrounding us and also infer from this light a wealth of information about the environment. We are, for instance, able to “see” the depth and shape of objects, the color of surfaces, or the segmentation of the scene into distinct objects. And we manage to do all this based just on images, i.e., two-dimensional distributions of intensities, which as such do not at all contain this kind of extra information. Computational vision and image analysis is a multidisciplinary scientific field that aims to give similar abilities to computers, i.e., to make them “see” in a way that is comparable to human perception. It is currently one of the most challenging research areas in artificial intelligence. By now it is widely recognized that mastering this scientific field is going to be one of the first key steps we must take towards achieving true artificial intelligence, which will require us to come up with answers to deep and fundamental questions about representation and computation lying at the core of human intelligence.

Essentially, computational vision can be viewed as some sort of information processing/extraction, where the goal is to process image data in order to extract a representation of objects in the world and to infer their properties. Note that the term image data in this case can have many different meanings depending on the context and the application at hand, e.g., it can refer to static photographs, video sequences, views from multiple cameras, X-ray data, infrared images, microscopy images etc. Similarly, the properties to be inferred from such data can refer to many different types of information such as depth, shape, surface color, object boundaries, object motion, to mention only a few of them. In fact, due to the proliferation of visual sensors as well as of storage devices that has taken place over the last years, we now observe an exponential growth of the stored visual content (e.g., images on the web, personal movies and photos, films, surveillance tapes, YouTube videos etc.).

As a result, the extraction of information from this vast amount of visual data and the exploitation of the resulting information space becomes an issue of even greater importance and remains one of the greatest challenges in our days. In other words, this means that there is a great need for general-purpose tools that will be able to effectively interpret the various types of visual data in an automated fashion. Addressing such a challenge has been one of the main focuses of my research. Broadly speaking, I am interested in using visual data such as natural images, video, or medical image modalities in order to “make sense of” (*i.e.*, analyze and understand) various aspects of the world that surrounds us. For instance, we,

as humans, simply open our eyes and seem to effortlessly recognize objects and the structures of scenes. But this apparent ease is highly misleading and reflects instead the enormous amount of neuronal resources (at least half the cortex) which is involved in performing these visual tasks. Moreover, in practice, the sheer volume of the available data makes impossible their analysis through human inspection alone.

To address the above challenges, it is therefore imperative to use mathematical models capable of explaining the huge amount of data available, and algorithms that can exploit the models to make predictions about the future. However, due to the tremendous ambiguities of the visual data and the enormous variability of the visual tasks, to be able to truly match the level of human performance in this area we need models that are truly rich and powerful. At the very least, such models must be probabilistic in nature to be able to account for the uncertainty present in the visual data. More importantly, however, such models must be capable to encode the rich dependencies/relationships existing in the input visual signal and to reason globally about the visual scene. Effective computational tools for handling/manipulating models of this type are therefore of paramount importance, and play an essential role for helping us to achieve further progress in this area. Thus, an important tenet of my scientific research work so far has been that:

One of the main scientific challenges for achieving true breakthroughs in the semantic interpretation of visual data relates to our ability in building and utilizing richer, more holistic models for the underlying visual tasks.

Motivated by such a need, over the past years I have focused my research efforts around the following fundamental tasks:

1. Inference for visual perception, where I have tried to introduce a very general computational framework based on which one will be able to perform efficient inference for an extremely broad class of models related to tasks in computer vision and image understanding.
2. Learning of models for visual perception, where I have also tried to introduce novel machine learning algorithms for automatically estimating the structure and the parameters of such models based on training data, thus aiming to address some of the key challenges related to visual learning tasks at the same time.
3. Developing of rich and efficient models for visual perception, where my goal has been to fully utilize the aforementioned inference and learning frameworks in order to propose novel concrete models for a variety of important

problems from the domains of computer vision, image processing and medical image analysis.

My motivation behind such an effort has been to help in further advancing the available technology for automated interpretation of visual data, which can in turn have a great societal and economic impact. Note, for instance, that advances in scene understanding can in turn bring significant improvements in areas as diverse as autonomous robots, surveillance, navigation & driver safety, situated search, industrial inspection, medical image analysis, content-based image retrieval and multimedia indexing, and intelligent vehicle systems, just to name a few of the affected areas.

Of course, in addressing the 3 aforementioned fundamental tasks has led me to utilize tools, combine results, and make contributions to a number of different areas, including computer vision & image analysis, machine learning, and discrete optimization. Moreover, besides the above areas of research, due to the fundamental and generic nature of the inference and learning methods that I have developed, these are also easily applicable to a number of other areas of artificial intelligence as well, such as, for instance, natural language processing, computational biology, data mining, pattern recognition, to name just a few of the possible application domains. Such an effort also directly relates to a second important tenet of my research, which is that

Developing a general inference and learning framework for visual perception based on graphical models not only is at the forefront of computer vision research today, but it will also have a profound impact in many other important branches of artificial intelligence.

This has been essentially one of the core topics addressed by my research work so far.

My goal in the following sections is exactly to provide a high-level overview of all this work conducted by me after obtaining my PhD. This overview has been roughly divided into two parts, comprising sections 2 and 3. The former section describes my research efforts towards developing a flexible, modular and computationally efficient inference and learning framework for image analysis. Such a framework has also formed the scientific foundations for the work presented in section 3. That section focuses on various applications of the aforementioned framework with respect to tackling fundamental and challenging problems from the fields of computer vision, image processing and medical imaging.

2 Scientific foundations

Developing computational solutions with respect to a specific visual task is a process that typically involves three components,

- (i) the parameterization of the problem through a set of parameters or a model,
- (ii) the association of this model with the available observations,
- (iii) and, the extraction of the optimal model parameters through an inference/optimization algorithm.

However, such a process presents several important challenges:

1. curse of non-linearity: often the observations are not directly associated with the model and therefore there is a non-linear relationship between them that makes inference quite challenging,
2. curse of non-convexity: in most of the cases the designed cost function is highly non-convex and therefore recovering computationally the optimal solution is not obvious,
3. curse of dimensionality: many of the related optimization problems encountered in vision are of very large scale (involving, e.g., millions of variables). This is even more so nowadays due to the great ease with which huge amount of visual data can be collected. As a result, computational efficiency is an extremely important factor.

On top of the above challenges, the models/algorithms that will be used must be able to face the immense variability of the visual tasks encountered in areas like computer vision and image analysis. As such, they must be robust enough and should also rely upon solid mathematical principles. Furthermore, they should be able to account for uncertainty, which is ever-present in the visual data (e.g., due to noise, imperfect sensors, or ambiguities in the visual interpretation). Discrete graphical models like Conditional Random Fields (CRFs) or Markov Random Fields¹ (MRFs) comprise a very elegant framework that satisfies all of the above properties. They provide a useful abstraction for quantifying uncertainty, describing complex dependencies in data while making the model's structure explicit so that it can be exploited by algorithms [53]. As such, they can and have been used for expressing a wide array of problems in computer vision [10, 120]. This explains why the tasks of performing inference and learning the parameters of these models are considered of paramount importance and have attracted a tremendous

¹The terms MRFs and CRFs will be used interchangeably throughout.

amount of research in the computer vision and machine learning communities over the past 40 years or so [26, 86, 99, 12, 36, 41, 135, 56, 119, 48, 49, 128, 123, 122]. Yet, as already explained above, such tasks are highly non-trivial (e.g. inference in such models is in general NP-hard).

To fully specify a graphical model of this class, one needs to provide a graph $G = (V, E)$, consisting of a set of nodes V and a set of edges E , as well as a set of so-called unary potentials $\mathbf{u} = \{u_p(\cdot)\}_{p \in V}$ and pairwise potentials $\mathbf{v} = \{v_{pq}(\cdot)\}_{pq \in E}$, where all these potentials are typically assumed to be parameterized through a vector of parameters \mathbf{w} , i.e., it holds

$$u_p(x_p) = u_p(x_p; \mathbf{w}), \quad v_{pq}(x_p, x_q) = v_{pq}(x_p, x_q; \mathbf{w}).$$

The energy of the resulting MRF model is then given by

$$\text{MRF}_G(\mathbf{x}|\mathbf{u}, \mathbf{v}) := \sum_{p \in V} u_p(x_p) + \sum_{pq \in E} v_{pq}(x_p, x_q) , \quad (1)$$

where $x_p \in L$ denotes the value assigned to the variable corresponding to node p in the graph, and L denotes the set of possible values/labels for that node. Such a model offers great flexibility and representational power, since by making a proper choice of its main elements, i.e., the graph, the labels and the MRF potential functions, one can express a very wide range of problems from image analysis and beyond.

Given such a model, the goal of inference is to estimate the minimum of the above energy function, which will hereafter be denoted by $\text{MRF}_G(\mathbf{u}, \mathbf{v})$ and is equal to

$$\text{MRF}_G(\mathbf{u}, \mathbf{v}) := \min_{\mathbf{x}} \text{MRF}_G(\mathbf{x}|\mathbf{u}, \mathbf{v}) . \quad (2)$$

The goal of learning, on the other hand, is to make use of some available training data in order to estimate the correct values that should be assigned to the parameters \mathbf{w} of the model so as to faithfully represent a specific task at hand.

Over the past years, a significant part of my research has been devoted to developing novel, highly practical and accurate inference/learning algorithms for such models, mainly targeting applications in computer vision and image analysis. One of the main goals of such an effort has been to introduce a modular and computationally efficient inference framework for visual image perception to address the most fundamental problems in computational vision. In the following I briefly present some of the main aspects of such a framework, thereby also describing some of the main themes of my research in this regard along with some of the long standing research challenges they aim to address.

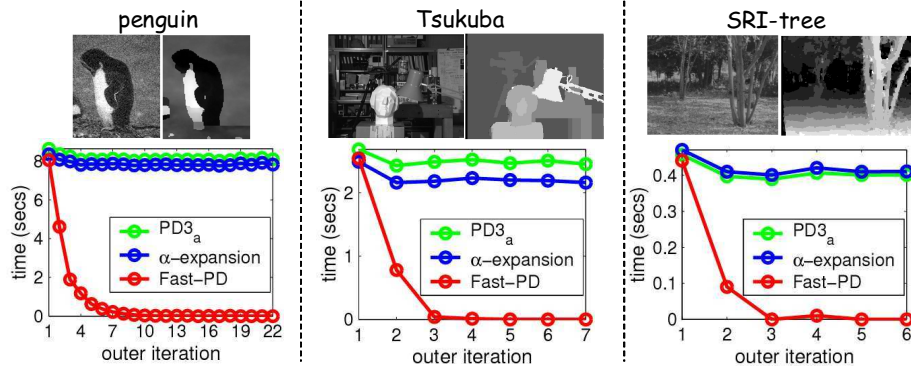


Fig. 1: FastPD running times compared to other MRF inference algorithms for some typical vision problems.



Fig. 2: Online approximation factors computed by FastPD for a benchmark stereo matching problem. Notice how quickly these approximation factors approach the value of 1, meaning that the corresponding estimated solutions are almost optimal.

2.1 Efficient inference on graphical models based on the duality theory of linear programming

In [69], [70], I have developed a very general framework for inference based on the primal-dual schema for linear programming. Such a schema is a well known technique in combinatorial optimization, as it has been used both for deriving exact polynomial-time algorithms to many cornerstone problems in combinatorial optimization (including max-flow, matching, shortest path, minimum branching and minimum spanning tree [98]) and for providing powerful approximation algorithms to many NP-hard combinatorial problems (such as those of set-cover, steiner-network, scheduling, steiner tree, feedback vertex set, just to mention a few examples [125, 40]). The primal-dual schema has been first introduced into computer vision by Komodakis *et al* [64, 65]. In our recent work [69, 70] we have built upon and extended that previous work to provide a framework that makes the following important contributions compared to prior art:

Computational efficiency for single MRFs: State of the art graph-cut based optimization algorithms for MRFs, such as the α -expansion method [11], try to optimize the MRF energy by solving a series of max-flow problems. Their efficiency is thus largely determined from the efficiency of these max-

flow problems, which, in turn, depends on the number of augmenting paths per max-flow. By building upon the framework proposed in [65], we have been able to derive a significantly improved primal-dual MRF optimization method, called Fast-PD [69]. This method, like [65] or α -expansion [12], also ends up solving a max-flow problem for a series of graphs. However, unlike these techniques, the graphs constructed by Fast-PD ensure that the number of augmentations per max-flow decreases dramatically over time, thus boosting the efficiency of MRF inference (see Fig. 1). To show this, we prove a generalized relationship between the number of augmentations and the so-called *primal-dual gap* associated with the original MRF problem and its dual. Furthermore, to fully exploit the above property, we have also proposed two new extensions: an *adapted max-flow algorithm*, as well as an *incremental graph construction* method.

Accuracy of solutions: Despite its efficiency, the proposed method also makes no compromise regarding either the quality of the solutions it generates or the generality of the MRFs it can handle. So, for instance, if the pairwise potentials $v_{pq}(\cdot, \cdot)$ are assumed to be metric functions, then it can be proved that Fast-PD is as powerful as α -expansion, in the sense that it computes exactly the same solution, but with a substantial speedup. Moreover, it applies to a much wider class of MRFs², *i.e.*, it can even handle MRFs for which the pairwise potentials $v_{pq}(\cdot, \cdot)$ are non-metric functions. In fact, in all these cases, the proposed method can provide theoretical (*i.e.*, worst-case) upper bounds about how far the energy of the generated solution can be from the unknown optimal MRF energy. Moreover, besides these theoretical upper bounds, our method is also capable of providing per-instance upper bounds that are also updated online, *i.e.*, during the executing of the algorithm. In practice, these bounds prove, of course, to be much tighter (*i.e.*, much closer to 1) than the worst-case upper bounds and hence can be very useful for assessing how well the algorithm has performed for a particular task at hand (see Fig. 2).

Efficiency for dynamic MRFs: Furthermore, besides being able to significantly speed up the optimization of static MRFs, Fast-PD can also be used for boosting the efficiency of dynamic MRFs, *i.e.*, MRFs whose parameters may change over time. This is an important class of models, which are often encountered in computer vision applications (e.g., when analyzing time varying signals such as video data). Two works that have been proposed in this regard recently are [52, 45]. These methods can be applied to dynamic MRFs that are binary or have convex priors. On the contrary, Fast-PD naturally

²Fast-PD requires only $v_{pq}(a, b) \geq 0$, $v_{pq}(a, b) = 0 \Rightarrow a = b$

handles a much wider class of dynamic MRF models. It manages to achieve that by also exploiting information coming from a problem that is dual to the original MRF optimization problem. Fast-PD can thus be thought of as a generalization of previous techniques. As a result of its great efficiency and accuracy, FastPD turns out to be able to provide approximately optimal solutions to dynamic NP-hard models even in real time, thus opening the way for a wide range of new applications in this regard.

2.2 MRF energy minimization and beyond via dual decomposition

In [63], [61], I have also introduced a new rigorous theoretical framework to address discrete MRF-based optimization in computer vision. Such a framework exploits the powerful technique of Dual Decomposition. In particular, it is based on a projected subgradient scheme [113, 91, 92, 80] that attempts to solve an arbitrary MRF optimization problem by first decomposing it into a set of appropriately chosen easy-to-handle MRF subproblems and then combining their solutions in a principled way. By analyzing the very weak conditions that these subproblems have to satisfy, I have been able to show that such an approach provides extreme generality and flexibility, thus leading to a very elegant framework that allows for designing powerful MAP estimation algorithms for a very wide class of problems.

Based on this framework one is thus able to derive message-passing techniques that, on the one hand, generalize and provide new insights into existing state-of-the-art approaches such as tree-reweighted methods [128, 54, 55], and, on the other hand, also enjoy much better theoretical properties at the same time.

By using, for instance, as slave subproblems any set of spanning trees that cover the MRF graph G , one can show that such an approach leads to efficient optimization schemes [61] that provably solve the following widely used LP relaxation to MRF optimization problem (2) (also known as the local marginal polytope relaxation), which lies at the heart of most state of the art MRF inference methods [134, 16, 108, 130, 76]:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \sum_{p \in V} \mathbf{u}_p \cdot \mathbf{x}_p + \sum_{pq \in E} \mathbf{v}_{pq} \cdot \mathbf{x}_{pq} \\ \text{s.t.} \quad & \mathbf{x} \in \mathcal{X}^G \end{aligned} \tag{3}$$

where the set \mathcal{X}^G is defined for any MRF graph $G = (V, E)$ and discrete label set L as follows:

$$\mathcal{X}^G = \left\{ \mathbf{x} \left| \begin{array}{ll} \sum_{l \in L} x_p(l) = 1, & \forall p \in V \\ \sum_{l' \in L} x_{pq}(l, l') = x_p(l), & \forall pq \in E, \forall l \in L \\ x_p(\cdot) \in \{0, 1\}, & \forall p \in V \\ x_{pq}(\cdot, \cdot) \in \{0, 1\}, & \forall pq \in E \end{array} \right. \right\}$$

More generally, this framework provides great freedom as to how the slave subproblems are chosen for a given inference problem $\text{MRF}_G(\mathbf{u}, \mathbf{v})$ [63]. For instance, assuming a decomposition of the graph $G = (V, C)$ into a set of subgraphs $\{G_i = (V_i, C_i)\}$ such that $V = \cup V_i$, $C = \cup C_i$, the following set of slaves $\{\text{MRF}_{G_i}(\mathbf{u}^i, \mathbf{v}^i)\}$ (that are MRFs defined on the subgraphs G_i having their own potentials $\mathbf{u}^i, \mathbf{v}^i$) leads to a convex dual relaxation of the following form:

$$\text{DUAL}_{\{G_i\}}(\mathbf{u}, \mathbf{v}) = \max_{\{\mathbf{u}^i\}, \{\mathbf{v}^i\}} \sum_i \text{MRF}_{G_i}(\mathbf{u}^i, \mathbf{v}^i) \quad (4)$$

$$\text{s.t. } \sum_i \mathbf{u}^i = \mathbf{u} \quad (5)$$

$$\sum_i \mathbf{v}^i = \mathbf{v} \quad (6)$$

where conditions (5), (6) simply express the fact that the sum of the potentials of the slaves should give back the potentials of the master MRF model.

In this manner, simply by choosing different decompositions $\{G_i\}$, different algorithms can be derived via the above scheme, all of which can be shown to provably optimize (possibly different) dual relaxations to the MRF inference problem. In each case, the sum of the minimum energies of the slaves always provides a lower bound to the minimum energy of the master MRF, and the maximum of these bounds coincides with the optimum of the underlying dual relaxation. As a result, appropriately chosen slave MRFs (*e.g.*, with more complex topology for the subgraphs G_i) can lead to better lower bounds and thus to more powerful underlying dual relaxations.

Moreover, in this manner, one is given the opportunity to derive inference techniques that can take full advantage of the special structure that may exist in any particular class of MRFs, which is one of the most important advantages of the above approach and also allows the use of efficient inference techniques such as, *e.g.*, graph-cut based methods.

2.3 Tighter LP relaxations and cycle repairing

As already mentioned in an earlier section, the local marginal polytope LP relaxation [16, 108] lies at the heart, and is thus closely connected to, the great majority of the state of art MRF inference techniques. Despite their success, however, all these LP-based methods are doomed to fail if the aforementioned relaxation does not approximate well the actual MRF inference problem (*i.e.*, it is not tight), which is exactly what happens in several cases where one has to deal with very hard MRF problem instances.

Motivated by such an observation, in [58] I have tried to specifically focus my

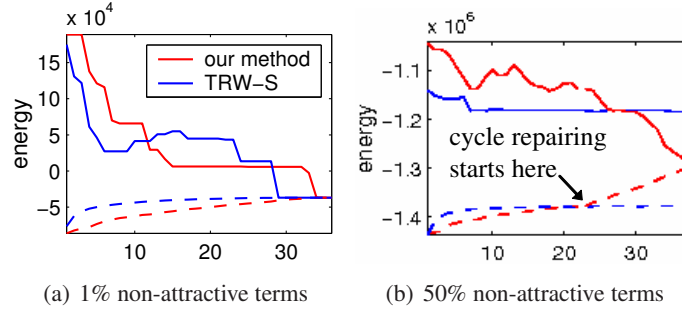


Fig. 3: (a) Upper (solid) and lower (dashed) bounds after running TRW-S and our method on 2 different binary MRFs. Such bounds correspond to primal costs (*i.e.*, MRF energies) and dual costs respectively. For a binary MRF with only 1% non-attractive terms, the upper and lower bounds of TRW-S converge (*i.e.*, the standard local marginal polytope LP-relaxation is tight). (b) This is not the case, however, for a binary MRF with 50% non-attractive terms. But, thanks to cycle repairing, our algorithm uses a much tighter relaxation and so its upper and lower bounds converge even in this case.

attention on MRF problems where the relaxation is known to be loose (*i.e.*, the solution of the relaxed problem is not optimal for the original one), trying to make both practical and theoretical contributions in this regard. In particular, the focus of my work has been to attempt to go beyond existing MRF inference techniques, by deriving algorithms that are based on LP relaxations that are much tighter than the aforementioned local marginal polytope relaxation. But instead of attempting to do that in the primal domain, which would have been presumably inefficient, my strategy has been to apply this tightening procedure in the dual domain. As a result of this strategy, a hierarchy of tighter and tighter dual relaxations is created that starts from the dual of the local marginal polytope relaxation and goes all the way up to a dual relaxation that is actually tight, *i.e.*, it coincides with the original MRF inference problem. From this hierarchy, we choose to deal with one particular class of relaxations, which we call *cycle-relaxations*, that turn out to provide the best trade-off between computational efficiency and tightness of approximation. This is achieved via an efficient dual-based operation called *cycle-repairing*, which helps us to better deal with a difficulty that lies at the core of why MRF optimization is actually an NP-hard problem: the existence of *inconsistent* cycles. As the name of that operation reveals, its role is to eliminate any inconsistent cycles that may appear during optimization. Furthermore, the more the repaired cycles, the tighter the underlying relaxation becomes.

It should be noted at this point that there have also been other works such as [115] that have recently tried as well to make use of tighter LP relaxations in the context of MRF optimization. However, the aforementioned method relies on a weaker relaxation than ours. Furthermore, they use a primal-based cutting plane

algorithm that requires solving a large primal LP (of growing size) at each iteration (i.e., after each new violated inequality is found), which makes their algorithm impractical for large scale MRF instances. On the contrary, by working in the dual domain, a method like ours is able to improve the relaxation (i.e., perform cycle repairing) by reusing work done in previous iterations. It is thus much more efficient, while it is also adaptive as it makes the relaxation tighter only when it needs to be. Moreover, being dual-based, it can provide lower bounds to the optimum MRF energy, which can be useful for verifying/assessing a solution's optimality. Other recent works that have focused on this important issue of tightening the underlying MRF relaxation are [76, 131, 114, 105, 47, 46].

In order to briefly describe the used hierarchy of dual relaxations, one needs to start with the dual to the local marginal polytope relaxation, which is the basic building block and apparently lies at one end of this hierarchy. For an MRF with unary and pairwise potentials $\bar{\mathbf{g}}, \bar{\mathbf{f}}$, this dual relaxation will hereafter be denoted by $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$. The dual cost of any feasible solution to that relaxation is a lower bound to the unknown minimum energy $\text{MRF}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$. However, it is often the case that even the maximum of these bounds will be much lower than the optimum MRF energy, which is exactly what happens when $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ is not tight.

To counter that, *i.e.*, to raise the maximum lower bound, one can resort to the relaxation $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ lying at the other end of our hierarchy, defined as follows for an MRF on a graph $G = (\mathcal{V}, \mathcal{E})$ having potentials $\bar{\mathbf{g}}, \bar{\mathbf{f}}$:

$$\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}}) = \max_{\mathbf{f}} \mathcal{D}(\bar{\mathbf{g}}, \mathbf{f}) \quad (7)$$

$$\text{s.t. } \mathbf{f} \preceq_{\mathcal{E}} \bar{\mathbf{f}}. \quad (8)$$

In (8), we have used an abstract comparison operation $\preceq_{\mathcal{E}}$ between pairwise potential functions. In general, given any subset of edges $\mathcal{C} \subseteq \mathcal{E}$ edges and any pairwise potential functions \mathbf{f}, \mathbf{f}' , the operation $\mathbf{f} \preceq_{\mathcal{C}} \mathbf{f}'$ means that the following inequality should hold true for any labeling $\mathbf{l} = \{l_p\}$:

$$\sum_{pp' \in \mathcal{C}} f_{pp'}(l_p, l_{p'}) \leq \sum_{pp' \in \mathcal{C}} \bar{f}_{pp'}(l_p, l_{p'}), \quad \forall \mathbf{l} = \{l_p\} \quad (9)$$

That is, instead of comparing the values of pairwise potentials on individual edges, we compare the sums of pairwise potentials values over all edges belonging to the set \mathcal{C} .

The reason that we have expressed $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ in the above form is in order to better illustrate its relation to relaxation $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$. As can be observed, one difference between $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ and $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ is that the latter contains an additional set of variables $\mathbf{f} = \{f_{pp'}(\cdot, \cdot)\}$, which can actually be thought of as a new set of pairwise

potentials (also called *virtual potentials* hereafter). In relaxation $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$, these virtual potentials \mathbf{f} do not appear at all, as they are essentially kept fixed to the true potentials $\bar{\mathbf{f}}$ (*i.e.*, it is as if constraints (8) have been replaced with the trivial constraints $\mathbf{f} = \bar{\mathbf{f}}$). On the contrary, in $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$, we can vary these new potentials in order to achieve a higher dual objective value (*i.e.*, a higher lower bound to the optimum MRF energy). The restriction, of course, is that we must never allow \mathbf{f} to become larger than the actual potentials $\bar{\mathbf{f}}$, where the comparison between \mathbf{f} and $\bar{\mathbf{f}}$ is done based not on the standard operator \leq , but on the generalized operator $\preceq_{\mathcal{E}}$. As a result of this fact, relaxation $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ can actually be shown to be tight.

One thus can argue that relaxations $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$, $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ lie at opposite ends: $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ imposes trivial constraints on \mathbf{f} and is thus efficient but not tight, whereas $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ has an exponential number of constraints on \mathbf{f} and is thus not easy to handle but is actually tight. By going, however, between these two ends, one can adjust the amount of constraints on \mathbf{f} through concentrating only on the virtual potentials at a subset of edges $\mathcal{C} \subseteq \mathcal{E}$ of the MRF graphedges. Indeed, assuming that initially $\mathbf{f} = \bar{\mathbf{f}}$, and that all $f_{pp'}(\cdot, \cdot)$ with $pp' \notin \mathcal{C}$ will be kept fixed during the current step, constraints (8) then reduce to the easier upper-bounding constraints $\mathbf{f} \preceq_{\mathcal{C}} \bar{\mathbf{f}}$, thus creating a relaxation in between $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ and $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$. Contrary to $\mathbf{f} \preceq_{\mathcal{E}} \bar{\mathbf{f}}$, constraints $\mathbf{f} \preceq_{\mathcal{C}} \bar{\mathbf{f}}$ focus only on a subset of the virtual potentials $f_{pp'}(\cdot, \cdot)$, *i.e.*, only on those with pp' in subset \mathcal{C} , while the rest are left untouched. Not only that, but, as optimization proceeds, one can choose a different local subset \mathcal{C}_i to focus on at each step. In this manner, different constraints can be *dynamically* used at each step, implicitly creating a dual relaxation that becomes tighter and tighter as time passes by. Such a relaxation is actually part of an hierarchy of dual relaxations that starts from $\mathcal{D}(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ (*e.g.*, if each \mathcal{C}_i contains only a single edge) and goes all the way up to $\mathcal{D}^+(\bar{\mathbf{g}}, \bar{\mathbf{f}})$ (*e.g.*, if \mathcal{C}_i contains all possible MRF edges). *Cycle-relaxations* refer to one particular type of relaxations from this hierarchy, where the edges from each set \mathcal{C}_i are always assumed to form a simple cycle on the MRF graph, and so, in this case, the process of enforcing the constraints associated with the dynamically chosen cycles is referred to as cycle repairing (Fig. 3).

2.4 Accelerating MRF inference

Message passing methods are among the most popular MRF optimization techniques in computer vision, with BP being the earliest method of this kind. Recently, many state of the art message-passing techniques based on dual decomposition have been proposed that rely on solving dual LP relaxations [61, 63, 128, 54, 130]. Compared to BP, they offer significant advantages such as better convergence properties, as well as the ability to provide suboptimality guarantees based on dual lower bounds. Moreover, they have been shown to significantly outperform BP

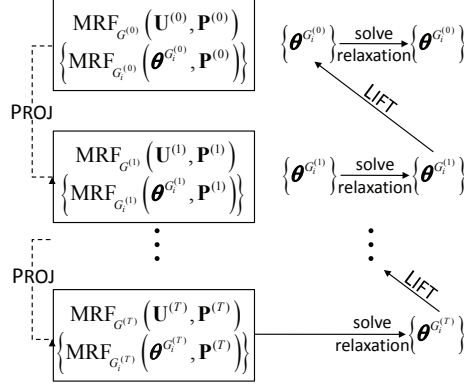


Fig. 4: V-cycle of the algebraic multigrid approach for dual LP-based algorithms

and other MAP estimation techniques [120]. On the other hand, one main drawback is that they often have a higher computational cost compared, *e.g.*, to graph-cut based methods [70, 2]. As a result, given the large scale nature of the majority of vision problems, one of the key challenges in energy minimization is how to be able to significantly accelerate this type of techniques. This is even more so considering the fact that computer vision researchers start gradually resorting to the use of higher order MRF models, where such dual-based methods are expected to have much wider applicability due to their generality.

Motivated by the above observations, in one of my recent works [57] I have focused on this central issue, *i.e.*, on how to increase the overall efficiency of dual LP-based algorithms, while maintaining or even improving their effectiveness (*i.e.*, their accuracy) at the same time. To that end, I have proposed a framework that integrates together two very general techniques in order to significantly speed up such algorithms. The first one is inspired by *algebraic multigrid* techniques for linear systems of equations, and uses a multiresolution hierarchy of dual relaxations for accelerating the convergence of dual-LP based methods. It relies on the premise that information is expected to propagate faster at lower resolutions. In the past, a geometric multigrid approach has been used for accelerating the BP algorithm, but is applicable only to grid-structured graphs [20]. On the contrary, I extended and generalized such an approach to LP-based algorithms, where a novel algebraic multigrid framework is able to handle MRFs defined on any kind of graph, or having any kind of potentials. Moreover, it can be applied to LP relaxations that are tighter than the standard marginal polytope relaxation.

The general idea in this case is to use a hierarchy of dual decompositions, defined on a sequence of graphs $G = G^{(0)}, G^{(1)}, \dots, G^{(T)}$, where each graph $G^{(t+1)}$ can be thought of as a “coarser” version of graph $G^{(t)}$. Due to the decomposition of the master MRF into a set of smaller slave MRFs, the update of the dual vari-

ables is essentially done based only on local information. As a result, information travels slowly across the graph, and this has the undesirable effect of slowing down the convergence of dual LP-based algorithms, which thus require many iterations to converge to the correct solution. This issue is essentially very similar to the slow convergence problem faced by iterative algorithms for linear systems. Again, due to the local nature of the updates, such algorithms can recover very fast (*i.e.*, in few iterations) the high-frequency part of the solution, but they are very slow at recovering the lower frequencies. Multigrid is introduced to overcome this problem, where the basic idea is based on the trivial observation that low frequencies in the original grid reappear as high frequencies in a grid of lower resolution. A multigrid approach thus replaces the original linear system with a hierarchical multiresolution set of linear systems. The two key elements in a multigrid algorithm are the so called restriction and prolongation operators, that specify the transition between linear systems at adjacent levels in the hierarchy. These operators are combined to generate a so called V-cycle, which consists of a fine-to-coarse restriction phase followed by a coarse-to-fine prolongation phase.

Therefore, in our approach we also need to define a restriction and prolongation operator, denoted hereafter by PROJ and LIFT respectively. The role of the restriction operator is to take as input a master MRF and its dual decomposition at level t , and to project them onto level $t + 1$, *i.e.*, to create a corresponding master problem and a corresponding dual decomposition at level $t + 1$

$$\begin{aligned} \text{MRF}_{G^{(t)}}(\mathbf{U}^{(t)}, \mathbf{P}^{(t)}) & \xrightarrow{\text{PROJ}} \text{MRF}_{G^{(t+1)}}(\mathbf{U}^{(t+1)}, \mathbf{P}^{(t+1)}) \\ \{\text{MRF}_{G_i^{(t)}}(\boldsymbol{\theta}^{G_i^{(t)}}, \mathbf{P}^{(t)})\} & \xrightarrow{\text{PROJ}} \{\text{MRF}_{G_i^{(t+1)}}(\boldsymbol{\theta}^{G_i^{(t+1)}}, \mathbf{P}^{(t+1)})\} \end{aligned} \quad (10)$$

On the contrary, the role of the prolongation operator LIFT is to take as input a feasible set of dual variables $\{\boldsymbol{\theta}^{G_i^{(t+1)}}\}$ for the decomposition defined at the “coarser” level $t + 1$, and to lift them to a feasible set of dual variables $\{\boldsymbol{\theta}^{G_i^{(t)}}\}$ for the decomposition that has been previously defined at level t , *i.e.*,

$$\{\boldsymbol{\theta}^{G_i^{(t+1)}}\} \xrightarrow{\text{LIFT}} \{\boldsymbol{\theta}^{G_i^{(t)}}\} . \quad (11)$$

Just like in multigrid, a V-cycle in our case will consist of a restriction phase followed by a prolongation phase (see Fig. 4). In the restriction phase we sequentially apply operator PROJ to all but the last level in the hierarchy, *i.e.*, we start from level $t = 0$ and go up to level $t = T - 1$. In this manner, a master MRF along with a dual decomposition is generated for each level. All of these decompositions are essentially projections of the original master problem and its dual decomposition. In the prolongation phase, we move in the opposite direction. This means that for each level t (where t now starts from $t = T$ and terminates at $t = 0$) we solve the

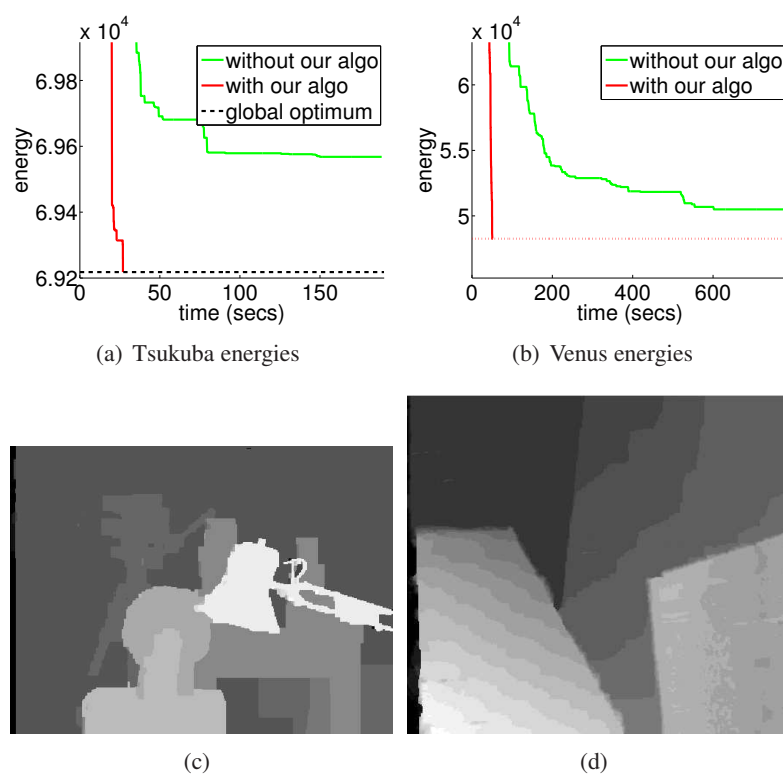


Fig. 5: Convergence plots for Tsukuba and Venus with and without our MRF inference acceleration method [57], as well as corresponding stereo matching results of our method.

dual relaxation corresponding to the decomposition at that level, and then we lift the resulting solution onto the next finer level (if one exists) via using the operator LIFT, thus initializing the dual variables for the decomposition at level $t - 1$. Due to the the information traveling much faster at the “coarser” levels of the hierarchy, the dual relaxations for these levels can be solved very fast, *i.e.*, in very few iterations. Furthermore, this quick spreading of the information that took place in the coarser levels is carried over to the finer levels, thanks to the initialization of the dual variables via the LIFT operator. This, in turn, results into accelerating the convergence of the dual relaxations at the finer levels as well.

Last but not least, besides the use of the above algebraic multigrid approach for accelerating the convergence of MRF inference, one also needs to significantly reduce the time per iteration of a dual LP-based algorithm in order to achieve an even greater speed up. Towards that goal, I have also introduced a second technique, which consists of a novel decimation strategy that carefully fixes the labels for a growing subset of nodes during the course of the algorithm, thus eliminating the need to update their dual variables thereafter. It is based on the observation that, when using an algebraic multigrid approach, a set of nodes typically exists that contribute a very small increase to the objective of the dual relaxation when their dual variables are updated. Similarly to the first technique, it is very general, and thus applicable to a very broad class of MRFs. Furthermore, it allows better primal solutions to be computed. Note that MRF decimation techniques have also been used in the past, and have been applied either to variants of BP [66, 13] or to dual LP-based algorithms [75, 2, 111]. However, the latter techniques are not as widely applicable as our method.

All in all, the use of the two aforementioned techniques that I proposed enables dual LP-based algorithms for MRF inference to improve both their speed of convergence and also the accuracy of their estimated solutions at the same time (e.g., see Fig. 5).

2.5 Towards efficient inference for high-order graphical models

With a few exceptions only, most of the existing inference techniques currently used in computer vision are confined to the case of graphical models containing low-rank (e.g., pair-wise) interactions between variables. One reason for this fact is because optimization of higher order MRFs can often be extremely challenging (*i.e.*, algorithms that yield almost optimal solutions are hard to get in this case) and, furthermore, these algorithms often have a very high computational cost that is prohibitive in practice. Yet, it is by now widely recognized that many vision problems could greatly benefit from the use of higher order models as this would allow far more expressive priors to be encoded, and also multiple interactions to be

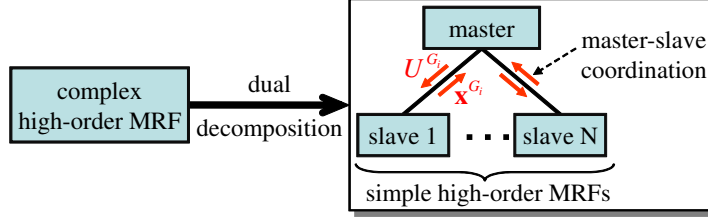


Fig. 6: Master-slave decomposition for inference on MRFs with high-arity factors.

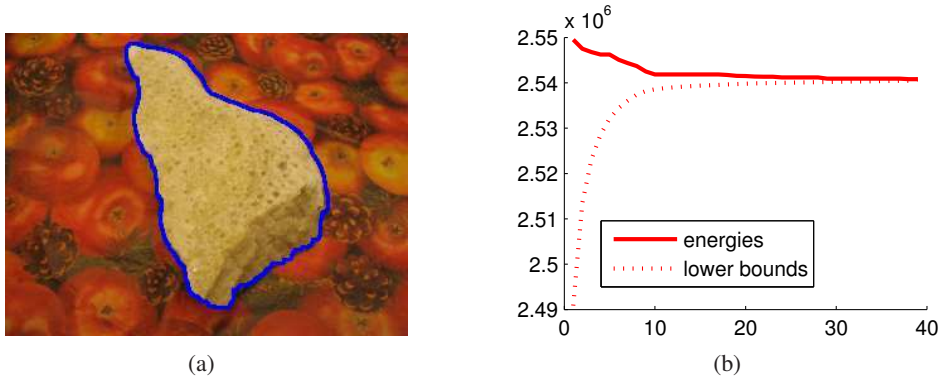


Fig. 7: (a) A binary image segmentation result computed by applying our high-order inference algorithm [59] to a $\mathcal{P}^{3 \times 3}$ Potts model. (b) In this case our method computes the global optimum of the corresponding 9th-order MRF since the MRF energies and the lower bounds finally become equal to each other. Note that, in the above plot, solid lines represent MRF energies, whereas dashed lines represent dual costs, i.e., lower bounds on the optimum MRF energy.

captured [103, 79, 100]. This would lead, in several cases, to a far better and more accurate modelling, which is clearly required in many vision tasks (such a need is also evident from the fact that in a variety of cases there is a large disagreement between the global optimum, that can often be computed for pairwise MRFs, and the ground truth solution [88]). In general, the energy of such MRF models is given by

$$E_G(\mathbf{x}) = \sum_{p \in V} u_p(x_p) + \sum_{e \in \mathcal{E}} \phi_e(\mathbf{x}_e),$$

where now \mathcal{E} denotes a set of cliques or hyperedges³, and $\phi_e(\mathbf{x}_e)$ (with $\mathbf{x}_e = \{x_p | p \in e\}$) denotes the corresponding higher-order potentials defined over these hyperedges.

Towards dealing with the above mentioned issues, I have recently proposed a powerful framework for efficient high-order MRF optimization [59]. It uses a

³A hyperedge is simply a subset of MRF nodes

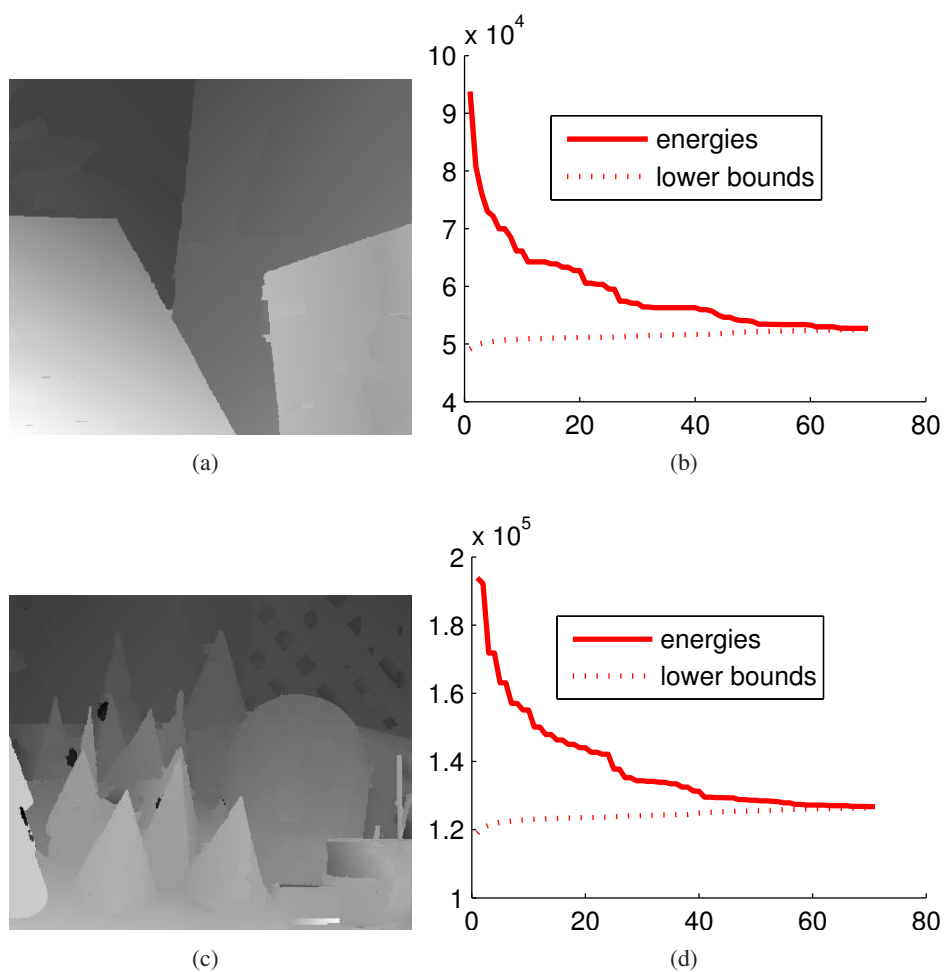


Fig. 8: Stereo matching results for 'venus' and 'teddy' (from the middlebury dataset) computed based on a higher-order model that uses a discontinuity preserving smoothness prior with second order derivatives. The plots show the corresponding energies and lower bounds during MRF energy minimization by our method.

master-slave based scheme (Fig. 6), which relies on the core idea that even a hard high-order MRF problem (with, *e.g.*, large cliques or complicated structure) can often be decomposed into high-order MRF subproblems that are very easy or even trivial to solve. This leads to a very general and flexible framework. In particular, such a framework can, on the one hand, be used for deriving a generic optimizer, which is applicable to almost any high-order MRF, and which provably computes the global optimum to a strong dual LP-relaxation to the MRF energy minimization problem.

On the other hand, due to its flexibility, the proposed framework can also be easily adapted to lead to even more powerful algorithms when it comes to dealing with specific classes of high-order MRFs. To further illustrate this, I have also introduced a new class of high-order potentials, called *pattern-based* potentials, which offer great expressive power and can be useful for a variety of computer vision tasks. By relying again on the same framework, a powerful and extremely efficient message-passing algorithm is proposed to handle this class of high-order potentials. This algorithm goes beyond the aforementioned generic optimizer and is able to deliver solutions of very high quality. As a result, for the first time, we have been able to show experimentally that in many practical cases one can compute the global optimum for NP-hard high-order MRFs used in vision, and, furthermore, we can do that in a very efficient manner, *e.g.*, at a fraction of the time that would be required by a generic message-passing scheme (Figures 7, 8). In this manner, our goal has been, through our framework, to further promote the applicability of higher-order models to vision.

It should be noted that the number of efficient MRF inference algorithms that have been proposed for dealing with high-order vision problems is quite sparse. A notable exception is the recent work of Kohli *et al* [51, 50], where an efficient inference technique was proposed for a specific class of higher-order MRFs. Lan *et al* [79] presented an efficient but approximate version of BP, while Potetz [100] proposed a BP adaptation for a certain class of high-order graphical models. The n -ary max-sum diffusion method has been very recently proposed by Werner [132], while two other works [42, 104] address high-order MRF optimization by reducing it to a pairwise problem with binary or multi-label variables, which is shown to lead to a compact representation in certain special cases.

2.6 Learning of high-order graphical models

In areas such as computational vision and image analysis, it is very often the case that hand-crafted models cannot cope with the complexity of the encountered problems. In many cases, for instance, the models that have to be employed and utilized may very well depend on a large number of parameters, and so trying to manually

tweak the values of these parameters can be a hopeless or (in the best case) an extremely laborious process. As a result, besides inference, the task of training a graphical model (i.e., determining its parameters/structure so as to provide an accurate representation of the problem at hand) plays an equally important role for successfully applying MRFs to problems from the above domains (it is not surprising, for instance, that a MAP-MRF solution is of little value if the used MRF does not properly represent the problem at hand).

Yet, such a task is highly non-trivial and presents significant challenges. This is because, unlike standard machine learning tasks where one must learn functions predicting simple true-false answers or scalar values (as in classification and regression), here one is supposed to learn models that predict answers much more complex that consist of multiple interrelated variables (it is a characteristic example of a so-called structured prediction problem).

Furthermore, this difficulty becomes even greater due to the computational challenges that are often raised by computer vision applications with regard to learning. For instance, many of the MRFs used in vision are of large scale. Also, the complexity and diversity of vision tasks often require the training of MRFs with complex potential functions. On top of that, during the last years the use of high order MRFs is becoming increasingly popular in vision since such models are often found to considerably improve the quality of estimated solutions. Yet, most of the MRF learning methods proposed so far in the vision literature compromise with regard to at least one or more of the above issues. For instance, most of these methods impose restrictions on the type of the MRF potential functions that can be used during learning, and/or can handle only pairwise MRFs [3, 77, 107, 121, 90, 89, 2].

In general, for training a graphical model, the provided input is assumed to consist of a set of K training samples $\{\mathbf{z}^k, \mathbf{x}^k\}_{k=1}^K$, where \mathbf{z}^k and \mathbf{x}^k represent respectively the observed data and the ground truth MRF label assignments of the k -th sample. Moreover, it is assumed that the unary potentials u_p^k and the pairwise potentials v_c^k of the k -th MRF training instance can be expressed linearly in terms of feature vectors extracted from the observed data \mathbf{z}^k , that is, it holds $u_p^k(x_p) = \mathbf{w}^T g_p(x_p, \mathbf{z}^k)$, $v_c^k(\mathbf{x}_c) = \mathbf{w}^T g_c(\mathbf{x}_c, \mathbf{z}^k)$, where $g_p(\cdot, \cdot)$ and $g_c(\cdot, \cdot)$ represent some known vector-valued feature functions (which are chosen based on the computer vision application at hand) and \mathbf{w} is an unknown vector of parameters. The goal of MRF training is exactly to estimate this vector \mathbf{w} using as input the above training data.

Both generative (e.g., maximum-likelihood) [77] and discriminative (e.g., max-margin) [122, 25] MRF learning approaches have been proposed in the literature for this purpose. In the former case, one seeks to maximize (possibly along with an L_2 norm regularization term) the product of posterior probabilities of the ground

truth MRF labelings given by

$$\prod_k P_G(\mathbf{x}^k; \mathbf{w}),$$

where $P_G(\mathbf{x}; \mathbf{w}) \propto \exp(E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}))$ denotes the probability distribution induced by an MRF model with energy $E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w})$ (recall that the notation $E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w})$ refers to the energy of an MRF defined on a graph G having some unary and higher-order potentials \mathbf{u}, \mathbf{v} both of which are assumed to be parameterized by \mathbf{w}).

This leads to a convex differentiable objective function that can be optimized using gradient ascent. However, computing the gradient of this function involves taking expectations of the feature functions $g_p(\cdot), g_c(\cdot)$ with respect to the MRF distribution $P(\mathbf{x}; \mathbf{w})$. One therefore needs to perform probabilistic MRF inference, which is, in general, an intractable task. As a result, approximate inference techniques (e.g., loopy belief propagation) are often used for approximating the MRF marginals required for the estimation of the gradient (which is suboptimal and prevents a thorough theoretical analysis of such techniques). This is the case, for instance, in [107], where the authors demonstrate how to train a CRF model for stereo matching, as well as in [77], where also a comparison with other CRF training methods such as pseudo-likelihood and MCMC-based contrastive divergence are included.

In the case of max-margin learning [123, 90], on the other hand, one seeks to adjust the vector \mathbf{w} such that the energy of the desired ground truth solution \mathbf{x}^k is smaller by $\Delta(\mathbf{x}, \mathbf{x}^k)$ than the energy of any other solution \mathbf{x} , that is,

$$E_G(\mathbf{x}^k; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) \leq E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) - \Delta(\mathbf{x}, \mathbf{x}^k) + \xi_k, \quad (12)$$

where $E_G(\cdot)$ denotes the energy function of an MRF defined on a graph G whose potentials \mathbf{u}, \mathbf{v} are parameterized by \mathbf{w} . In the above set of linear inequality constraints with respect to \mathbf{w} , $\Delta(\mathbf{x}, \mathbf{x}')$ represents a user-specified distance function (such as the Hamming distance) that measures the dissimilarity between any two solutions \mathbf{x} and \mathbf{x}' (obviously it should hold $\Delta(\mathbf{x}, \mathbf{x}) = 0$). Furthermore, ξ_k is a non-negative slack variable introduced for ensuring that a feasible solution \mathbf{w} always exists. Ideally, \mathbf{w} should be set such that each variable $\xi_k \geq 0$ can take a value as small as possible (so that, in effect, the amount of total violation of the above constraints is minimal). As a result, during learning the following constrained op-

timization problem has to be solved

$$\begin{aligned} \min_{\mathbf{w}, \{\xi_k\}} \quad & \mu \cdot R(\mathbf{w}) + \sum_{k=1}^K \xi_k \\ \text{s.t. constraints (12)} \end{aligned} \quad (13)$$

In the above problem, μ is a user-specified hyperparameter, and $R(\mathbf{w})$ represents a regularization term whose role is to prevent overfitting during the learning process (e.g., it can be set equal to $\|\mathbf{w}\|^2$ or to a sparsity inducing norm such as $\|\mathbf{w}\|_1$). The slack variable ξ_k can also be expressed as the following hinge-loss term:

$$\mathcal{L}_G(\mathbf{x}^k; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) = E_G(\mathbf{x}^k; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) - \min_{\mathbf{x}} \left(E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) - \Delta(\mathbf{x}, \mathbf{x}^k) \right). \quad (14)$$

This, in turn, leads to the following equivalent unconstrained formulation:

$$\min_{\mathbf{w}} \mu \cdot R(\mathbf{w}) + \sum_{k=1}^K \mathcal{L}_G(\mathbf{x}^k; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}). \quad (15)$$

One class of methods [23, 87] try to solve the constrained optimization problem (13) by use of a cutting-plane approach when $R(\mathbf{w}) = \|\mathbf{w}\|^2$. In this case, the above problem is equivalent to a convex quadratic program (QP) but with an exponential number of linear inequality constraints. Given that only a small fraction of them will be active at an optimal solution, cutting plane methods proceed by solving a small QP with a growing number of constraints at each iteration (where this number is polynomially upper-bounded). One drawback of such an approach relates to the fact that computing a violated constraint requires solving at each iteration a MAP inference problem that is NP-hard in general. For the special case where the MRF potentials are constrained to be submodular, [3] show how to express the above constraints (12) in a compact form, which allows for a more efficient MRF training to take place in this particular case.

Towards addressing all aforementioned challenges, in a recent strand of my work I have focused on this important topic of graphical model learning, where I have proposed a novel max-margin framework specifically for that purpose [71]. Such a framework makes use of recent advances made on the MRF optimization side [74, 73], which are, in this case, combined for the first time with a max-margin approach for MRF training [123]. In particular, the dual decomposition approach [74], which has been previously used for MAP estimation, is now employed for this purpose as well.

Note, in this regard, that the main difficulty for minimizing functional (15) stems from the intractability of the term $\min_{\mathbf{x}} (E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) - \Delta(\mathbf{x}, \mathbf{x}^k))$

that is included in the definition of the loss $\mathcal{L}_G(\mathbf{x}^k; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w})$, which is NP-hard to compute in general. Assuming without loss of generality that $\Delta(\mathbf{x}, \mathbf{x}^k) = \sum_p \delta(x_p, x_p^k) + \sum_c \delta(x_c, \mathbf{x}_c^k)$, the above intractable term can be equivalently expressed as

$$\min_{\mathbf{x}} \left(E_G(\mathbf{x}; \mathbf{u}^k, \mathbf{v}^k, \mathbf{w}) - \Delta(\mathbf{x}, \mathbf{x}^k) \right) = \min_{\mathbf{x}} E_G(\mathbf{x}; \bar{\mathbf{u}}^k, \bar{\mathbf{v}}^k, \mathbf{w}),$$

where $\bar{u}_p^k(\cdot) = u_p^k(\cdot) - \delta(\cdot, x_p^k)$, $\bar{v}_c^k(\cdot) = v_c^k(\cdot) - \delta(\cdot, \mathbf{x}_c^k)$ are the so-called *loss-augmented* MRF potentials. In order to deal with the intractability of the above minimization, I have proposed to resort to approximating it with a tractable convex relaxation. In particular, I have proposed to use convex relaxations derived from dual decomposition, which, in practice, have been previously shown to be quite tight, *i.e.*, it is assumed

$$\min_{\mathbf{x}} E_G(\mathbf{x}; \bar{\mathbf{u}}^k, \bar{\mathbf{v}}^k, \mathbf{w}) \approx \text{DUAL}_{\{G_i\}}(\bar{\mathbf{u}}^k, \bar{\mathbf{v}}^k; \mathbf{w}), \quad (16)$$

where $\{G_i\}$ can be any chosen decomposition of the graph G . Recently, I have been able to show [71] that through such an approximation, the learning problem (15) reduces to the following form

$$\min_{\mathbf{w}, \{\bar{\mathbf{u}}^{k,i}\}} R(\mathbf{w}) + \sum_k \sum_i \mathcal{L}_{G_i}(\mathbf{x}^k; \bar{\mathbf{u}}^{k,i}, \bar{\mathbf{v}}^k, \mathbf{w}) \quad (17)$$

$$\text{s.t.} \quad \sum_i \bar{\mathbf{u}}^{k,i} = \bar{\mathbf{u}}^k. \quad (18)$$

Essentially, one of the most important benefits from the above reduction is that it manages to convert the training of a complex MRF that is defined on a graph G (see the term $\mathcal{L}_G(\cdot)$ in (15)) to the parallel training of the slave MRF models defined on the subgraphs G_i (see the terms $\mathcal{L}_{G_i}(\cdot)$ in (17)), where the latter can be much easier to handle within a max-margin learning framework.

For instance, one can simply choose a decomposition where each subgraph G_i corresponds to exactly one clique of the original hypergraph G , which leads to slaves that are typically easy to train and thus to a very general learning scheme. On the other hand, there can also exist cases where such a decomposition may not provide the best possible result one can achieve both in terms of efficiency (*e.g.*, it can lead to slower convergence during training) but also in terms of accuracy. For instance, such a decomposition may not fully exploit additional properties that a given class of MRFs may have (it should be noted that such additional properties typically exist in many MRFs encountered in image analysis, *e.g.*, either due to the special form of the MRF potentials or due to the special topology of the graph G). However, the above learning framework can gracefully handle such cases as well,

since, thanks to it, one is still able to develop highly efficient and accurate learning schemes simply by using a more appropriate decomposition that is adapted to the model's structure while remaining tractable at the same time.

In all of the aforementioned cases, it is ensured that the concurrent training of the slave MRFs takes place in a principled way through an efficient projected subgradient algorithm. This leads to a powerful learning framework that makes the following important contributions compared to prior art: (1) it is able to efficiently handle not just pairwise but also high-order MRFs, (2) it does not impose any restrictions on the type of MRF potential functions that can handle or on the topology of the MRF graph, (3) the reduction to the parallel training of a series of slave MRFs in combination with the projected subgradient method [102, 6, 109] leads to a highly efficient learning scheme that is amenable to distributed computing and is also scalable even to very large problems, (4) it allows the use of a hierarchy of convex relaxations for approximating MAP-MRF estimation within learning for structured prediction (where this hierarchy is derived by using a series of decompositions and can be shown to include some widely used LP relaxations for MRF inference), thus leading to structured prediction learning algorithms of increasing accuracy, (5) last, but not least, it is extremely flexible and extendable since its only requirement to a user is to be able to compute an optimizer for a slave MRF, while everything else is taken care by the learning framework itself. As such, it can be easily adapted to take advantage of the special structure that may exist in any given class of MRFs that one wishes to train.

2.7 Learning of graphical models with weak supervision

When it comes to learning of graphical models for image analysis, another issue of utmost importance relates to the amount of available supervision. Most methods typically assume that training takes place using fully supervised data, i.e., each training sample specifies the ground truth values for all variables of the graphical model. However, as we transition into using more and more complicated models, the amount of time needed by a user to fully annotate a training example can vary substantially. For instance, although a simple model for object detection may only require specifying a bounding box for each training image, a more complex model might also require specifying a segmentation mask as well as bounding boxes for all parts of the object.

Given the large number of training images that are often available, it would be impractical to expect all of them to be fully annotated. Therefore, a well-designed learning method should be able to take full advantage of training examples with varying levels of annotation, i.e., both fully annotated data and data that are weakly annotated. On the one hand, the former type of data provide the greatest possible

amount of information but are typically sparse in number. On the other hand, it is very often the case that weakly annotated data (or even data with no annotation at all) are quite easy to obtain in massive quantities. Due to their large number, such data can therefore provide extremely valuable information and can guide the learning process despite the fact that they contain only partial information.

To address this very important issue, I have recently worked on developing novel learning schemes that can be used for performing an efficient training of so-called *latent graphical models* [72], which are models containing latent (i.e., hidden) variables that are not observable during both training and testing [136] (it should be noted that such models have lately been shown to play a crucial role in many important vision applications [19]).

Let E_G denote the energy of such a model defined on a hypergraph $G = (V, \mathcal{E})$ with vertices V and hyperedges \mathcal{E} . In this case, the energy function is defined as

$$E_G(\mathbf{x}, \mathbf{y}; \mathbf{w}) = \sum_{p \in V} u_p^k(x_p, y_p; \mathbf{w}) + \sum_{e \in \mathcal{E}} \phi_e^k(\mathbf{x}_e, \mathbf{y}_e; \mathbf{w}),$$

where we now assume that \mathbf{x}, \mathbf{y} represent respectively the observed and hidden variables of the model. Furthermore, functions $u_p^k(\cdot)$, $\phi_e^k(\cdot)$ denote the unary and higher-order potentials that are again assumed to be expressible in terms of an unknown vector of parameters \mathbf{w} and some feature functions of the input data \mathbf{z}^k (e.g., $u_p^k(x_p, y_p; \mathbf{w}) = \mathbf{w}^T f_p(x_p, y_p, \mathbf{z}^k)$ and similarly for $\phi_e^k(\cdot)$). As in the case of supervised learning, we want to estimate \mathbf{w} based on a provided set of training samples $\{\mathbf{y}^k, \mathbf{z}^k\}_{k=1}^K$, with the important difference, however, that now only variables \mathbf{y}^k are observable during training (whereas variables \mathbf{x}^k remain hidden during both training and testing).

To deal with training models of the above type, I have recently proposed a discriminative learning framework [72] that relies on extending the dual-decomposition based approach used for fully supervised learning [71] to the above case. All in all, this results into deriving a very efficient iterative learning scheme that essentially keeps alternating between the following two main steps: (a) on the one hand, completing (in a principled manner) the values of all latent variables of the graphical model, (b) and, on the other hand, updating the parameters \mathbf{w} by training in parallel a set of fully supervised MRF slave subproblems.

2.8 Inference and learning for LP-based clustering

Clustering is considered among one of the most fundamental unsupervised learning tasks. It lies at the heart of many important problems in computer vision, image analysis and pattern recognition. Most of the clustering methods are center-based, thus trying to extract a set of cluster centers that best “describe” the input data.

Typically, this translates into an optimization problem where one seeks to assign each input data point to a unique cluster center such that the total sum of the corresponding distances is minimized. These techniques are extremely popular and they are thus essential even to other types of clustering algorithms such as Spectral Clustering methods [94, 126, 112].

In practice, however, there exist several issues that affect the performance and effectiveness of clustering. For instance, many center-based clustering methods require the input data points to have a vectorial form. However, such an assumption severely limits their applicability since often one wants to be able to deal with more complicated forms of data such as graphs or sets of varying cardinality. Moreover, in many cases an explicit data point representation may not even be available in the first place. Instead, data points may only be implicitly defined through the specification of a distance function between them.

A very important issue, in this case, is the type of distance functions (used for measuring dissimilarity between data points) that the algorithm can handle. Ideally, one would like to be able to cluster data based on arbitrary distances. This is so because the used distance function essentially determines the possible shape of a cluster (e.g., an Euclidean distance assumes that clusters have a spherical shape). Therefore, by an appropriate choice of these distances, clusterings with completely different characteristics can be realized [24].

Furthermore, the majority of center-based clustering methods rely on EM-like schemes for optimizing their clustering objective function [5]. K-means is the most characteristic (and perhaps the most widely used) technique from this class. It keeps greedily refining a current set of cluster centers based on a simple gradient descent scheme. As a result, it can very easily get trapped to bad local minima and is extremely sensitive to initialization. It is thus likely to fail in problems with, e.g., a large number of clusters.

Last, another crucial issue, which can have an important effect on the quality of clustering, relates to the correct number of clusters that needs to be extracted. Contrary to what actually holds in most real-world problems in image analysis and computer vision, the majority of clustering algorithms assume that this number is known a priori. Instead, the desired and most proper behavior is that the number of clusters should be also estimated automatically, e.g., as a byproduct of the optimization process.

Towards addressing the above challenges, I have recently worked on developing novel clustering techniques that aim to make important contributions regarding both inference and learning [62], [72]:

On the front of inference, I have proposed a novel center-based clustering

method, which utilizes the following exemplar-based clustering formulation

$$\min_{Q \subseteq S} E(Q) = \sum_{p \notin Q} \min_{q \in Q} d_{p,q} + \sum_{q \in Q} d_{q,q} . \quad (19)$$

In the above formula, S denotes a set of datapoints endowed with a distance \mathbf{d} , whereas Q represents the set of cluster centers (exemplars), which in this case can consist of any subset of data points from the input set S . The role of elements of $\mathbf{d} = \{d_{p,q}\}$ is twofold: for $p \neq q$ each element $d_{p,q}$ represents the distance between datapoints p and q , whereas each element $d_{q,q}$ represents the penalty for choosing q as exemplar. As a result of this formulation, one seeks to minimize the distance of a datapoint to its nearest center, while at the same time choosing as few centers as possible. It is important to note that, in this case, the number of cluster centers is not predetermined but is an output of the optimization.

The above NP-hard optimization problem can also be expressed as an equivalent linear integer program [15] for which we use linear programming and duality theory in order to compute an approximately optimal solution [62]. This leads to an efficient and very general algorithm, which works in the dual domain, and can cluster data based on an arbitrary set of distances. Despite its generality, it is independent of initialization (unlike EM-like methods such as K-means), has guaranteed convergence, is able to automatically determine the number of clusters, and can also provide online optimality guarantees about the quality of the estimated clustering solutions. The latter come in the form of lower bounds on the cost of the optimal clustering and are computed (for free) by simply using the cost of the dual solutions generated during the course of the algorithm.

To deal with the most critical issue in a center-based clustering algorithm (selection of cluster centers), we also introduce the notion of *stability* of a cluster center, which is a well defined LP-based quantity that plays a key role to the algorithm's success. Intuitively, the stability of a data point as a cluster center tries to measure how much one needs to penalize that point (by appropriately modifying the objective function) such that it can no longer be chosen as a center in an optimal solution of the modified problem. Apparently, one would like to choose as centers those points having high stability. To that end, I have also proposed a computationally efficient method for approximating the stabilities of the datapoints, which relies on properly utilizing available dual information.

Also, on the front of learning how to cluster, I have recently proposed a very general framework that can automatically estimate from training data the proper distance function that should be used for a given clustering task [72]. Due to the complexity and variability of the clustering problems encountered in computer vision, the ability to automatically learn such a distance function based on training

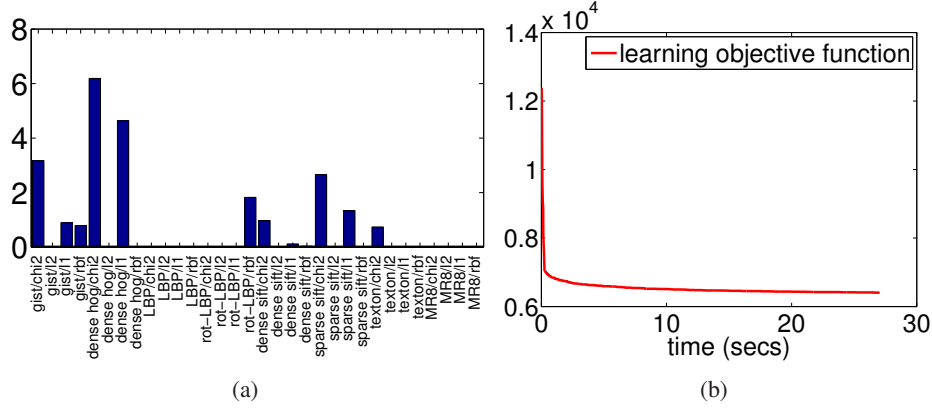


Fig. 9: (a) Learnt weights for the distance function for the 15 category scene dataset. In this case, the distance to be learnt is a weighted combination of individual distances, each one comparing images based on a different visual feature (various features are considered, including Gist [97], histogram of dense quantized HOG, local binary patterns (LBP) [96], rotation invariant LBP (rot-LBP) [1], histograms of quantized dense SIFT descriptors, histograms of quantized sparse SIFT descriptors (at Hessian-affine interest points), standard texon histograms and histograms based on MR8 textons [124]). (b) Evolution of the resulting learning objective function during training.



Fig. 10: 10 of the images that have been chosen as cluster centers for the Scene dataset.

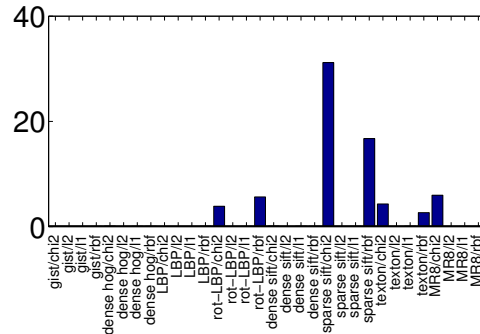


Fig. 11: Corresponding weights learnt for the distance function used in the UIUC texture dataset.

data is a matter of utmost importance for obtaining the best possible performance (moreover, in many cases this is actually the only viable choice as the distances to be used often depend on a large number of parameters, which precludes the possibility of manually adjusting all of them).

The proposed framework provides a very general max-margin learning scheme for distance-based clustering. It uses as input a set of ground truth partitions of training datasets. Based just on this input, it enables the discriminative learning of a very broad class of distances for clustering, where, *e.g.*, non-metric, non-differentiable or even non-symmetric distances can be handled by the proposed method. Despite its generality, the proposed learning framework provides great computational efficiency as it is based on a very fast projected subgradient method, and is, furthermore, inherently parallelizable. Also, by utilizing the formulation of clustering (19), it properly accounts for the fact that the number of clusters is typically not known in advance at test time, which is another important advantage compared to prior art.

In this case, the learning problem for clustering is formulated as one of training the following high-order conditional random field (CRF)

$$E(\mathbf{x}; \mathbf{d}) = \sum_{p,q} u_{pq}(x_{pq}; \mathbf{d}) + \sum_{p,q} \phi_{pq}(x_{pq}, x_{qq}) + \sum_p \phi_p(\mathbf{x}_p). \quad (20)$$

whose unary potentials $\mathbf{u} = \{u_{pq}(\cdot; \mathbf{d})\}$ and higher order potentials $\phi = \{\phi_{pq}(\cdot), \phi_p(\cdot)\}$ are defined as follows

$$u_{pq}(x_{pq}; \mathbf{d}) = d_{p,q} x_{pq} \quad (21)$$

$$\phi_{pq}(x_{pq}, x_{qq}) = \delta(x_{pq} \leq x_{qq}) \quad (22)$$

$$\phi_p(\mathbf{x}_p) = \delta\left(\sum_q x_{pq} = 1\right), \quad (23)$$

where $\mathbf{x}_p = \{x_{pq} | q \in S\}$, and $\delta(\cdot)$ equals 0 if the expression in parenthesis is satisfied and ∞ otherwise.

In the above MRF model, each binary variable x_{qq} indicates whether datapoint q has been chosen as a cluster center or not, and x_{pq} with $p \neq q$ indicates whether p has been assigned to the cluster with center q or not. The high order potentials $\phi_p(\mathbf{x}_p)$ in (23) ensure that each p is assigned to exactly one cluster, whereas the pairwise potentials $\phi_{pq}(x_{pq}, x_{qq})$ (22) ensure that if p is assigned to q then the latter must have necessarily be chosen as a cluster center.

Here, the distance function \mathbf{d} is assumed to be linearly parameterized by a vector of parameters \mathbf{w} . In this case, the two major challenges that one needs to deal with when trying to estimate this vector \mathbf{w} by training the above CRF model are, on the one hand, the fact that the above CRF contains factors of very high

order (e.g., note that the order of the potentials $\phi_p(\mathbf{x}_p)$ equals the cardinality of the set of datapoints S), and, on the other hand, the fact that the above CRF model contains latent variables that are not observable during training. To successfully handle both of these challenges, we rely on our previously mentioned learning framework for high-order latent graphical models, which is based on a master-slave dual decomposition approach. Figures 9, 10, 11 show two example applications of this training framework for clustering, which involve learning how to cluster scene images (e.g., 15 scene category dataset [82]) and how to cluster texture images (e.g., UIUC [81], Outex [95] datasets).

3 Applications to visual computing

Thanks to the generality of the aforementioned inference and learning techniques for visual image perception, I have been able to successfully apply them for providing state-of-the-art solutions to a wide variety of fundamental tasks from computer vision, image processing and medical image analysis. These include such problems as image-based 3D modeling of large natural environments, motion analysis and optical flow estimation, intelligent image completion & inpainting, tracking, automated texture synthesis, knowledge-based image segmentation, estimation of realistic 3D facial animations from video data, animal motion reconstruction from image sequences, deformable registration between images, segmentation and reconstruction of anatomical structures with prior knowledge, learning deformation priors for dense registration, diffusion tensor registration, image fusion through deformable mosaicing, manifold-based clustering, and group-wise (population) registration. In the following sections I briefly describe some of my contributions for a selective set of problems that I have worked on over the past years.

3.1 Computer vision and image processing

Blind image deconvolution using MRF-based image priors

Blind image deconvolution is a fundamental but very challenging problem, which has a long history in the image and signal processing literature [78]. Perhaps its most well known use is for removing the blur from consumer photographs (e.g. due to camera shake), but it also has important applications in areas such as computational photography and astronomical imaging [101, 85]. The input to this problem consists of a degraded image \mathbf{I} that equals the convolution of a true image \mathbf{x} with a kernel \mathbf{k} plus some noise \mathbf{n} , or

$$\mathbf{I} = \mathbf{x} \otimes \mathbf{k} + \mathbf{n}, \quad (24)$$

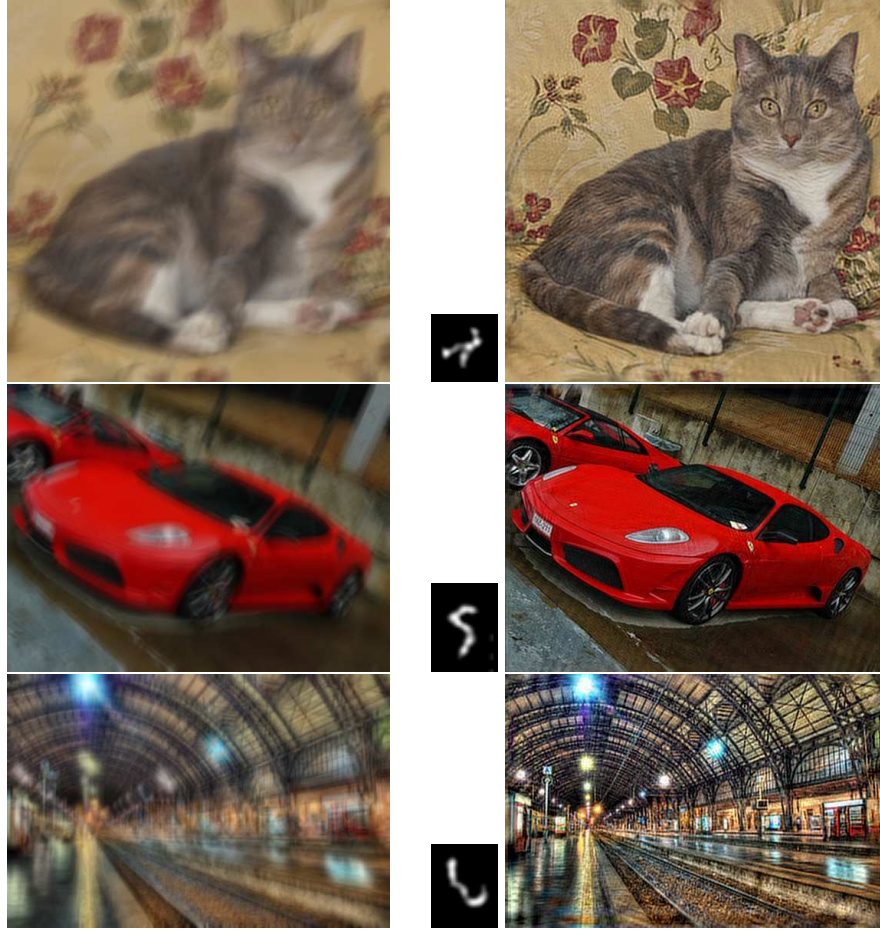


Fig. 12: **Left:** input images. **Right:** kernels k and deconvolved images x as computed by our method. The size (in pixels) of kernel k is, from top to bottom: 40×40 , 29×38 , 30×37 .

where \otimes denotes the convolution operator. Given only the image I as input, the goal of blind image deconvolution is to inverse the above process and to recover both x and k , which are assumed to be the unknowns in this case.

Over the past years, the problem of blind image deconvolution has attracted a significant amount of attention from the computer vision and image processing community, and thus a variety of algorithms [21, 110, 44, 17, 133, 4, 43, 84, 83] have been proposed that try to contribute to the state of the art in various ways. Obviously, one of the main difficulties of blind deconvolution relates to the fact that there can be exponentially many images x and kernels k that satisfy equation (24), which, in other words, means that inverting the above equation is a severely ill-posed problem [14].

To address this problem, I have recently proposed an optimization-based blind

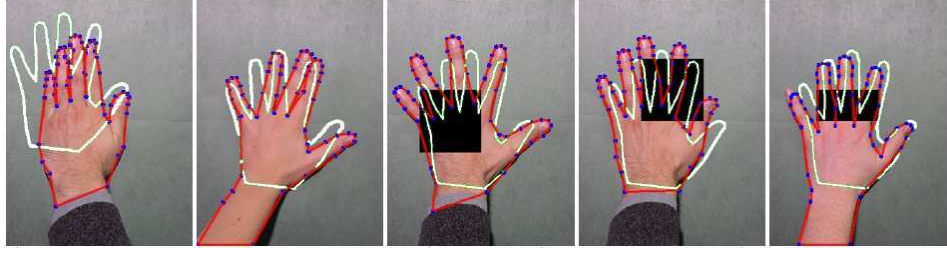


Fig. 13: The 10th frame of the Army sequence (upper row) and Mequon sequence (lower row) and our estimated (color-encoded) optical flow fields.

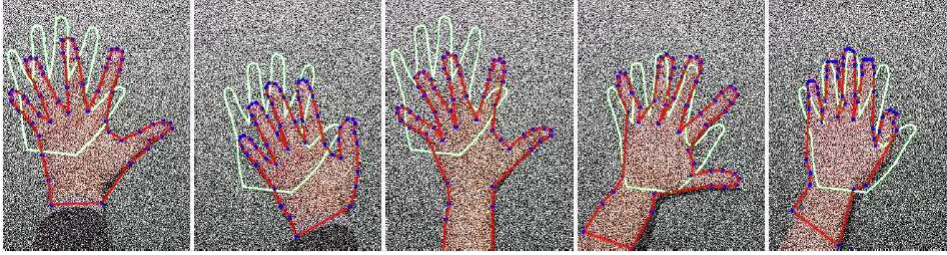
image deconvolution method that relies on imposing a discrete MRF prior on the deconvolved image [60]. The use of such a prior results in a very efficient and powerful deconvolution algorithm that carefully combines advanced optimization techniques such as fast inference methods for discrete MRFs [70] and the alternating direction method of multipliers [18, 22, 129]. Besides the computational efficiency of the proposed scheme, it has also been demonstrated that it can easily handle even very challenging blind deblurring problems that involve large and complicated blur kernels (Fig. 12).

Motion analysis and optical flow estimation

Optical flow estimation is a core task for the analysis and reconstruction of an object's motion, and is also heavily used in many other applications (e.g., video processing, super-resolution, noise reduction and removal etc.). It consists of recovering a 2D displacement vector establishing correspondences between the consecutive projections of a 3D patch in the image. Unfortunately, this is an inherently ill-posed problem, which presents great difficulties due to the ambiguities existing in the flow where different displacements locally might correspond to the same error (like in the absence of texture). In addition, another difficulty stems from the fact that in most cases it is absolutely critical that the displacement vectors are computed with very fine precision, i.e., with sub-pixel accuracy. To address these challenges, I have worked on novel discrete-to-continuous optical flow estimation and motion analysis techniques [33, 29, 7] that offer the following advantages com-



(a) Finger Collusion - Missing Part Examples. Two first images: difficult examples because of fingers collusions. Three last images: segmentation of hands with missing parts.



(b) Severe Noise Added: The prior knowledge highly contributes in correctly segmenting very noisy images.

Fig. 14: Model-based segmentation of the hand. Initialization is shown in white, segmentation in red, and the final control points positions in blue.

pared to prior art:

- (i) they have incrementally refined precision that is defined locally and varies according to the image structure,
- (ii) they can encode complex interactions between graph nodes,
- (iii) and last, but not least, can also complete the task in computationally efficient manner.

They rely on an energy minimization framework that automatically estimates uncertainty maps which are directly related with the covariance matrix of the obtained solution. These maps are used within a dynamic MRF model where the set of possible deformations is varying in space while being able to self-adjust the precision of the obtained optical flow vectors according to the observed uncertainties. They are thus able to compute optical flow fields of high quality (Fig. 13).

Knowledge-based image segmentation

Segmentation is a fundamental problem in computer vision and image processing. I have worked on a number of MRF-based approaches to address this task

[9, 38, 37]. An important strand of my work on this topic was concerned with how to exploit shape prior information in order to significantly improve the final accuracy and quality of the segmentation. To this end, I have proposed an image segmentation framework that makes proper use of such priors [9] along with clustering methods [62]. It includes novel representations to model geometric shape variations as well as efficient inference procedures to fit the resulting shape models to the input data. The considered shape model is similarity-invariant and refers to a sparse graph that consists of intra and inter-cluster connections representing the inter-dependencies of control points. The clusters are determined according to the correlations of the deformations of the control points within the training set using a novel clustering approach [62]. The connections between the components of a cluster represent the local structure while the connections between the clusters account for the global structure. The prior model is encoded as the distributions of the normalized distances between the connected control points. During search, this model is used together with a discrete Markov random field (MRF) based segmentation, where the unknown variables are the positions of the control points in the image domain. The resulting method is computationally efficient, can encode complex statistical models of shape variations, and is able to combine both local and global shape priors (Fig. 14). Moreover, due to its generality, it can be applied in a wide variety of different settings.

From images to 3D models: visual 3D reconstruction of large natural environments

An important research problem in computer vision is the creation of image-based modeling systems capable to provide photorealistic 3D representations of complex, real-world environments. Minimal human intervention during the modeling process, as well as operation in real time during rendering (a property that allows virtual walkthroughs at interactive frame rates) are some of the desirable characteristics for such systems. Towards that goal, I have proposed novel computer vision algorithms [68, 67] for a hybrid (geometry- and image-based) modeling and rendering framework that allows capturing real-world outdoor environments of very large scale and of complicated geometry. The only input required by such a framework is a sparse set of captured images from the scene. By applying advanced vision-based methods, a series of so-called morphable 3D-mosaics is automatically constructed from the captured images, which are then used for representing the entire scene. To this end, a continuous morphing between 3D-mosaics (that are nearby to the current viewpoint) is taking place during rendering. The morphing is both photometric and geometric. Moreover, due to the way that morphable 3D-mosaics have been constructed, this morphing is also ensured to proceed in

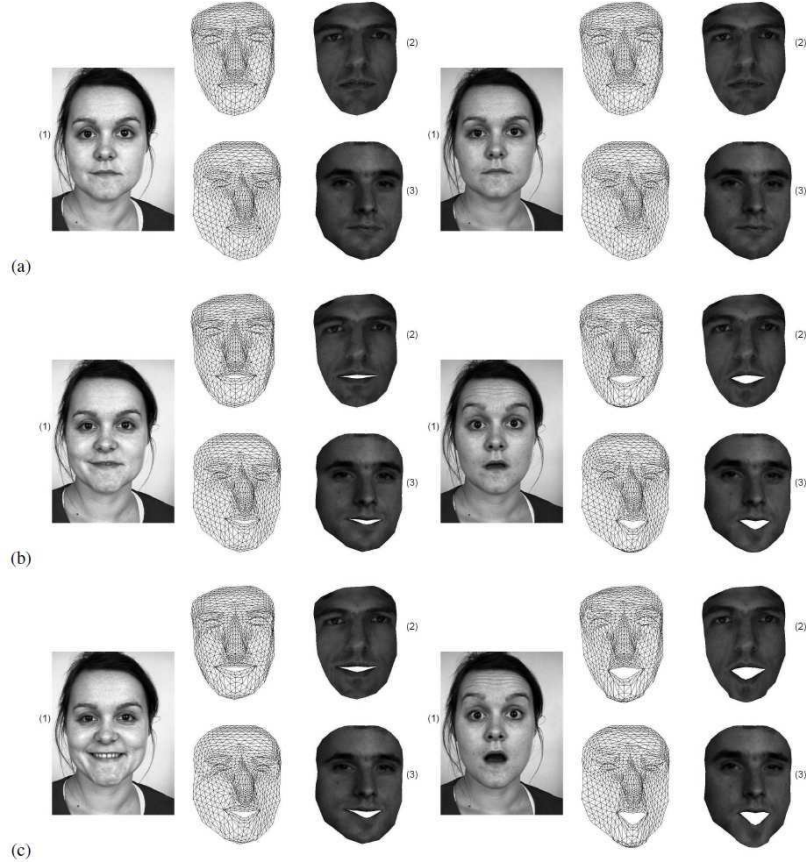


Fig. 15: Joy animation : (1) Observed sequence, (2) Individual 1, (3) Individual 2, (a) First frame , (b) intermediate frame, (c) last frame.

a physically valid manner, thus remaining transparent to the user. The proposed framework offers scalability to large scale environments, as only one morphable 3D mosaic needs to be displayed at any time during rendering. Furthermore, it can faithfully reproduce the photorealistic richness of a scene, while requiring minimal to none human intervention during the modeling process. On top of that, it runs at interactive frame rates thanks to using a rendering pipeline which is highly optimized for current 3D graphics hardware. Its effectiveness has been demonstrated in the automatic 3D visual reconstruction of the Samaria Gorge in Crete [68], which is one of the largest and most beautiful gorges in Europe.

Expression mimicking: from video to realistic 3D facial animations

Reproducing facial animations from images is a very challenging problem in computer vision. It can lead to highly realistic animation results and, as such, it has

received a substantial amount of effort with post-production and cinematography being the main driving forces of such innovation needs. I have recently worked [27] on a novel method for expression mimicking from a monocular video sequence to a 3D video avatar (Fig. 15). To this end, a generic 3D face mesh model is constructed first, through automatic global and local registration of low resolution range data. The expression mimicking problem is then addressed through a compact facial representation with control points (based on the MPEG-4 standard), and an efficient/optimal search of its geometric elements in the image. During the search procedure, weak classifiers along with cascaded Adaboost are used, while the optimal configuration of the Adaboost responses is found using discrete MRF optimization techniques based on linear programming, which enforce the anthropometric nature of the model. Last, but not least, animation is done using radial-based functions.

Tracking

Tracking is an essential part of many computer vision and image processing applications. I have recently proposed a novel end-to-end system for deformable tracking of multiple curvilinear objects in image sequences [39]. The approach is based on B-spline snakes defined by a set of control points whose optimal configuration is determined through efficient discrete optimization. Each control point is associated with a discrete random variable in a MAP-MRF formulation where a set of labels captures the deformation space. In such a context, generic terms are encoded within this MRF in the form of pairwise potentials. The use of pairwise potentials along with the B-spline representation offers nearly perfect approximation of the continuous domain. Efficient linear programming is considered to recover the approximate optimal solution. The proposed method performs in real-time, is shown to be robust to poor features and high deformations of the object to be tracked, can continuously maintain high accuracy on a sub-pixel level, and is generic in the choice of data and regularization terms. Moreover, the discrete framework can track multiple objects at the same time without altering or extending the model.

3.2 Medical image analysis

Medical image analysis is one of the most prominent application fields of computer vision and image processing. It is characterized by the extraction of information from image data for the purpose of making a medical diagnosis of a patient. This type of computer-aided diagnosis is increasingly considered in health sciences. This is due to the progress made on the acquisition side, where recent hardware developments have led to a new generation of scanners as well as image

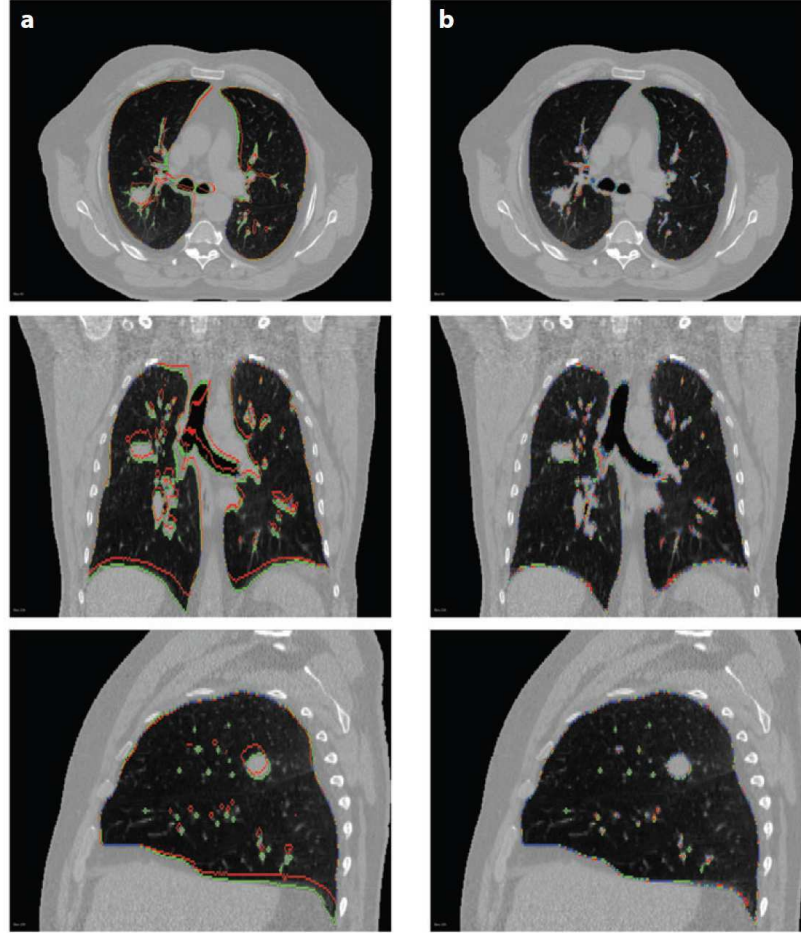


Fig. 16: Visual results (a) before and (b) after registration using our iconic approach [32]. Green voxels correspond to the segmentation of the reference; red voxels, to the (warped) segmentation. Voxels in blue show exact overlap of reference and warped segmentation.

modalities. As a result, in vivo visualization of human tissues where one can determine both anatomical and functional information is now possible. The analysis and processing of the resulting image data for the interpretation of the tissues state is a challenge of paramount importance that can significantly facilitate the task of the physicians. Registration, segmentation, and 3D reconstruction/modeling of anatomical structures are among the most fundamental problems in medical image processing that need to be solved for addressing such a challenge.

Deformable registration

The objective of deformable registration is to recover a deformation field that aligns two images (or volumes) that have in general an unknown non-linear relationship.

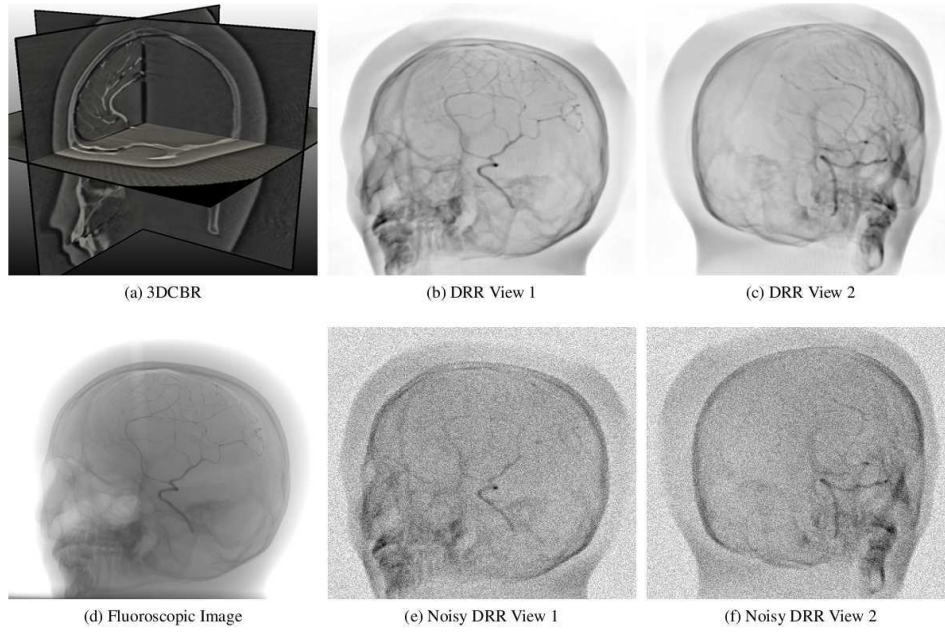


Fig. 17: Visualization of input data for the 2D-3D registration:(a) 3D cone beam reconstruction (3DCBR);(b,c) Exemplary digitally recon-structed radiographs (DRRs) which are computed in the iterations of the registration algorithm;(d) Fluoroscopic image used as target in 1-view test;(e,f)Exemplary DRRs with 20% uniform noise used as targets in 2-view tests.

This is a predominant task in medical image analysis involved in a vast number of clinical examples. Over the last few years I have developed several state-of-the-art techniques on this topic. These can be used for either inter or intra-modal deformable registration (Fig. 16) [32, 31], for hybrid (i.e., both iconic and geometric-based) registration [34], for intensity-based registration of images by linear transformations [138, 35], as well as for 2D-3D registration (Fig. 17) [137]. The proposed methods rely on expressing deformable registration as a minimal cost graph problem, where nodes correspond to the deformation grid, a node's connectivity corresponds to regularization constraints, and labels correspond to 2D/3D deformations. They lead to modular, powerful, and flexible MRF formulations that can account for arbitrary image-matching criteria, various local deformation models, and regularization constraints. Thanks to the use of powerful discrete optimization algorithms, they are able to attain unprecedented accuracy while requiring running times that are orders of magnitude lower compared to prior art.

Segmentation and reconstruction of anatomical structures

Organs segmentation and modeling are two fundamental tasks in medical image processing. The use of prior information either through explicit modeling of the anatomy or through learning carries on great potential. I have recently proposed general MRF-based methods to address these two tasks [8, 117, 106, 30], including methods that rely on using prior information of the above type. The proposed methods can deal with single or multi-component prior models, can encode parametric and non-parametric priors, and are able to use the entire information space. Among other things, I have applied them for the 3D shape modeling and segmentation of the left ventricle [8], for the automatic recovery of the 3D shape of the knee based on a set of 2D X-Ray images [106], as well as for cartilage segmentation using a statistical atlas [30].

Groupwise registration

Groupwise (population) registration is defined as the identification of a homology between more than two images. Its importance is evident in problems like statistical modeling of variations and atlas construction. I have recently proposed a novel registration framework that is able to unite a population of images to an optimal (unknown) pose through their mutual deformation [116]. It makes use of a registration criterion that comprises three terms: the first imposes compactness on appearance of the registered population at the pixel level, the second tries to minimize the individual distances between all possible pairs of images, while the last is a regularization one imposing smoothness on the deformation fields. The problem

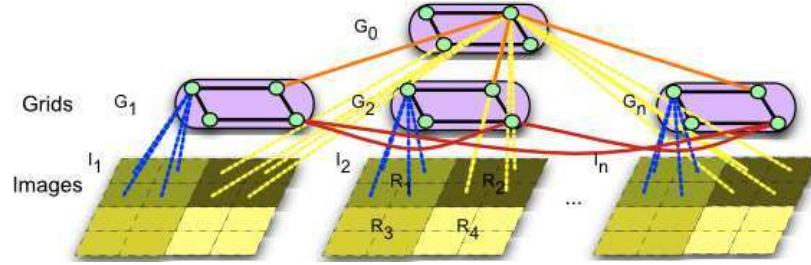


Fig. 18: The node and the edge system of the constructed graph. With blue color the relationship between the grid nodes and the images is depicted (deformation model). The black edges represent the smoothness terms while the red ones encode the local dissimilarity measure. The global relationship between all the nodes at respective places in the grids is shown by the yellow edges. (For clarity a fraction of the edges is shown.)

is reformulated as a graphical model (Fig. 18) that consists of hidden (deformation fields) and observed variables (intensities). A novel deformation grid-based scheme is proposed that guarantees the diffeomorphism of the deformation and is computationally favorably compared to standard deformation methods. Towards addressing important deformations, a compositional approach is also used where the deformations are recovered through approximately optimal solutions of successive discrete MRFs by using efficient linear programming.

Learning deformation priors for dense image registration

Learning appropriate prior models is of paramount importance to address the ill-posedness of the registration task, in particular when considering intra-modal registration of challenging, emerging imaging modalities such as functional MRI, diffusion tensor imaging, or ultrasound. Towards this goal, I have recently proposed a general framework for learning and inferring such priors based on image data [28]. The proposed priors can be used for replacing the conventional regularization constraints (e.g., penalizing the gradients of the displacement field) and can be very efficiently embedded within MRF-based registration algorithms. Moreover, they can be learned from a rather small training set, while they can encode local as well global prior constraints on the deformation field.

Image fusion through deformable mosaicing

Whole-body magnetic resonance imaging (WB-MRI) is an emerging application gaining vast clinical interest during the last years. It has been made clinically feasible thanks to recent advances in MRI such as multi-channel receiver, parallel imaging techniques, and automated table movement. Although such technological advances shortened the longish acquisition time, this is still the limiting factor

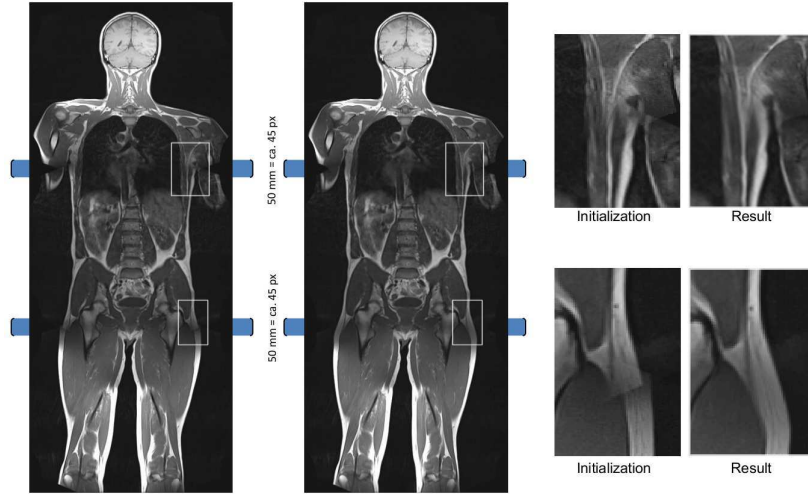


Fig. 19: Whole body MRI results (blue indicates overlap): (Left) initialization, (Middle) our result, (Right) zoomed-in image patches from the left and middle images.

avoiding its wide-spread clinical usage. The acquisition of images with large field-of-view helps to relieve this drawback, but leads to significantly distorted images. To address this critical issue, I have recently proposed a novel deformable mosaicing approach [127], based on the simultaneous registration to linear weighted averages, to correct for distortions in the overlapping area. This method produces good results on in-vivo data (Fig. 19) and has the advantage that a seamless integration into the clinical workflow is possible.

Processing of diffusion tensor images (DTI)

Registration: Diffusion tensor imaging is a fairly new modality that is able to provide clinicians with very useful information about the structure and the geometry of the observed tissues. In this case, diffusion tensor registration not only aims to recover the spatial correspondences but also reorient the tensors accordingly to account for the rotational component of the spatial deformation. The directional information of the diffusion tensors as well as the high-dimensionality of the data further complicates the registration process. To address these issues, I have recently proposed a novel method for the spatial normalization of diffusion tensor images [118]. This method takes advantage of both the diffusion information and the spatial location of tensor in order to define an appropriate metric in a probabilistic framework. A registration energy is defined in a reproducing kernel Hilbert space (RKHS), encoding the image dissimilarity and the regularity of the deformation field in both the translation and the rotation space. The problem is reformulated

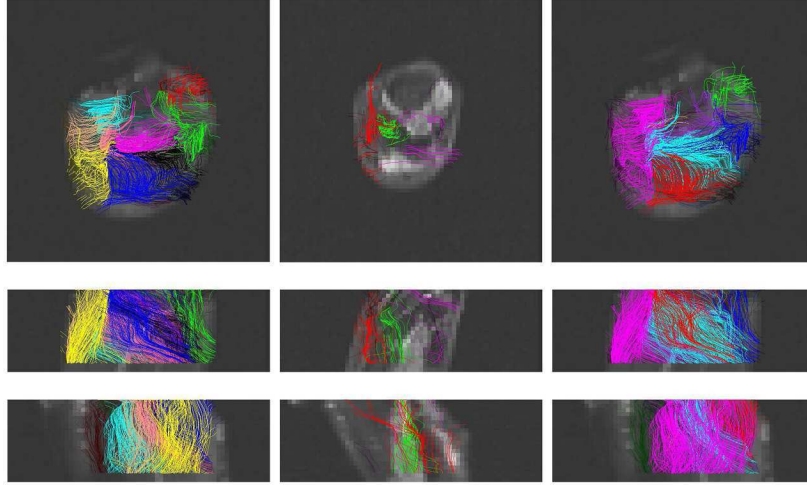


Fig. 20: Axial, coronal and sagittal views of fiber segmentation for (left column) a healthy subject in 10 classes (middle column) a diseased subject in 3 classes. In the right column it is shown the ground truth segmentation of the left column with the following muscles: the soleus (cyan), lateral gastrocnemius (red), medial gastrocnemius (magenta), posterior tibialis (yellow), anterior tibialis (green), extensor digitorum longus (purple), and the peroneus longus (blue).

as a graphical model where the latent variables are the rotation and the translation that should be applied to every tensor and the observed variables are the tensors themselves. Efficient linear programming is used to minimize the resulting energy.

Clustering: DTI has been mainly used to study the connectivity between the different structures of the human brain. Lately, it has also been used to study the human skeletal muscles as diffusion can provide information about the structure and the organization of the muscle fibers. A very important task in this regard is to cluster fiber tracts for local statistical analysis of diffusion information. Recently, I have proposed [93] a novel manifold-based clustering algorithm for this task (Fig. 20). Using a linear programming formulation of prototype-based clustering, I designed a novel fiber classification algorithm over manifolds that circumvents the necessity to embed the data in low dimensional spaces and determines automatically the number of clusters. I also proposed the use of angular Hilbertian metrics between multivariate normal distributions to define a family of distances between tensors that I generalize to fibers. These metrics are used for approximating the geodesic distances over the fiber manifold.

4 Acknowledgements

The results presented in this thesis would not have been made possible without the strong help and support of the many colleagues and students that I have collaborated with over the past years. I want, therefore, to express my sincere thanks and gratitude to each and everyone of them.

References

- [1] Timo Ahonen, Jivří Matas, Chu He, and Matti Pietikäinen. Rotation invariant image description with local binary pattern histogram fourier features. In *SCIA*, 2009.
- [2] Karteek Alahari, Chris Russell, and Philip Torr. Efficient piecewise learning for conditional random fields. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [3] Dragomir Anguelov, Ben Taskar, Vassil Chatalbashev, Daphne Koller, Dinkar Gupta, Jeremy Heitz, and Andrew Ng. Discriminative learning of markov random fields for segmentation of 3d scan data. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [4] S. Derin Babacan, Rafael Molina, and Aggelos K. Katsaggelos. Variational bayesian blind deconvolution using a total variation prior. *IEEE Trans. on Image Processing*, 18(1):12–26, January 2009.
- [5] Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. *J. Mach. Learn. Res.*, 6:1705–1749, 2005.
- [6] Dimitri Bertsekas. *Nonlinear Programming*. 1999.
- [7] A. Besbes, N. Komodakis, B. Glocker, G. Tziritas, and N. Paragios. 4D ventricular segmentation and wall motion estimation using discrete optimization. In *International Symposium in Visual Computing*, 2007.
- [8] A. Besbes, N. Komodakis, and N. Paragios. Graph-based knowledge-driven discrete segmentation of the left ventricle. In *IEEE International Symposium on Biomedical Imaging*, 2009.
- [9] A. Besbes, N. Komodakis, N. Paragios, and G. Langs. Shape priors and discrete MRFs for knowledge-based segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

REFERENCES

- [10] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov Random Fields for Vision and Image Processing*. MIT Press, 2011.
- [11] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11), 2001.
- [12] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Nov. 2001.
- [13] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: An algorithm for satisfiability. *Random Struct. Algorithms*, 27(2):201–226, 2005.
- [14] Patrizio Campisi and Karen Egiazarian. *Blind Image Deconvolution: Theory and Applications*. CRC Press, 2007.
- [15] Moses Charikar, Sudipto Guha, Éva Tardos, and David B. Shmoys. A constant-factor approximation algorithm for the k-median problem. *J. Comput. Syst. Sci.*, 65(1):129–149, 2002.
- [16] C. Chekuri, S. Khanna, J. Naor, and L. Zosin. Approximation algorithms for the metric labeling problem via a new linear programming formulation. In *12th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 109–118, 2001.
- [17] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *SIGGRAPH ASIA*, 2009.
- [18] Jonathan Eckstein and Dimitri P. Bertsekas. On the douglas-rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 1992.
- [19] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009.
- [20] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *Int. J. Computer Vision*, 70(1):41–54, 2006.
- [21] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. Removing camera shake from a single photograph. *Proc. ACM Siggraph*, 2006.

REFERENCES

- [22] Mario A.T. Figueiredo, Jose M. Bioucas-Dias, and Manya V. Afonso. Fast frame-based image deconvolution using variable splitting and constrained optimization. In *SSP*, 2009.
- [23] T. Finley and T. Joachims. Training structural svms when exact inference is intractable. In *Proc. Int. Conf. on Machine Learning*, 2008.
- [24] B. Fischer, V. Roth, and J. Buhmann. Clustering with the connectivity kernel. In *Advances in Neural Information Processing Systems*, 2004.
- [25] Vojtech Franc and Bogdan Savchynskyy. Discriminative learning of max-sum classifiers. *J. Machine Learning Research*, 2008.
- [26] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [27] C. Ghys, M.Taron, N. Paragios, N. Komodakis, and B. Bascle. Expression mimicking: from 2D monocular sequences to 3D animations. In *International Symposium in Visual Computing*, 2007.
- [28] B. Glocker, N. Komodakis, N. Navab, G. Tziritas, and N. Paragios. Dense registration with deformation priors. In *International Conference on Information Processing in Medical Imaging*, 2009.
- [29] B. Glocker, N. Komodakis, N. Paragios, and N. Navab. Approximated curvature penalty in non-rigid registration using pairwise MRFs. In *International Symposium in Visual Computing*, 2009.
- [30] B. Glocker, N. Komodakis, N. Paragios, N. Navab, G. Tziritas, and Ch. Glaser. Primal/dual linear programming and statistical atlases for cartilage segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, 2007.
- [31] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *International Conference on Information Processing in Medical Imaging*, 2007.
- [32] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N.Paragios. Dense image registration through MRFs and efficient linear programming. *Medical Image Analysis*, 2008.

REFERENCES

- [33] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab. Optical flow estimation with uncertainties through dynamic MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [34] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios. Deformable medical image registration: Setting the state of the art with discrete methods. *Annual Review of Biomedical Engineering (AR)*, 2011.
- [35] B. Glocker, D. Zikic, N. Komodakis, N. Paragios, and N. Navab. Linear image registration through MRF optimization. In *IEEE International Symposium on Biomedical Imaging*, 2009.
- [36] D. M. Greig, B. T. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society*, 1989.
- [37] I. Grinias, N. Komodakis, and G. Tziritas. Bayesian region growing and MRF-based minimization for texture and color segmentation. In *International Workshop on Image Analysis for Multimedia Interactive Services*, 2007.
- [38] I. Grinias, N. Komodakis, and G. Tziritas. Flooding and MRF-based algorithms for interactive segmentation. In *International Conference on Pattern Recognition*, 2010.
- [39] H. Heibel, B. Glocker, M. Groher, N. Paragios, N. Komodakis, and N. Navab. Discrete tracking of parametrized curves. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [40] Dorit S. Hochbaum, editor. *Approximation algorithms for NP-hard problems*. PWS Publishing Co., Boston, MA, USA, 1997.
- [41] H. Ishikawa. Exact optimization for markov random fields with convex priors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, October 2003.
- [42] Hiroshi Ishikawa. Higher-order clique reduction in binary graph cut. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [43] Neel Joshi, Richard Szeliski, and David J. Kriegman. Psf estimation using sharp edge prediction. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [44] Neel Joshi, C. Lawrence Zitnick, Richard Szeliski, and David J. Kriegman. Image deblurring and denoising using color priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

REFERENCES

- [45] Olivier Juan and Yuri Boykov. Active graph cuts. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [46] Fredrik Kahl and Petter Strandmark. Generalized roof duality for pseudo-boolean optimization. In *International Conference on Computer Vision*, 2011.
- [47] J. Kappes, S. Schmidt, and C. Schnörr. MRF inference by k-fan decomposition and tight Lagrangian relaxation. *Proc. ECCV*, pages 735–747, 2010.
- [48] J.H. Kappes, B. Savchynskyy, and C. Schnörr. A Bundle Approach To Efficient MAP-Inference by Lagrangian Relaxation. In *Proc. CVPR*, 2012.
- [49] J.H. Kappes, M. Speth, B. Andres, G. Reinelt, and C. Schnörr. Globally Optimal Image Partitioning by Multicuts. In *Proc. EMMCVPR*, 2011.
- [50] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [51] Pushmeet Kohli, Pawan Kumar, and Philip Torr. P3 and beyond: Solving energies with higher order cliques. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [52] Pushmeet Kohli and Philip H.S. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *International Conference on Computer Vision*, 2005.
- [53] D. Koller and N. Friedman. *Probabilistic graphical models: Principles and techniques*. MIT Press, 2009.
- [54] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006.
- [55] Vladimir Kolmogorov and Martin Wainwright. On the optimality of tree-reweighted max-product message passing. In *Conference on Uncertainty in Artificial Intelligence*, 2005.
- [56] Vladimir Kolmogorov and Ramin Zabih. What energy functions can be minimized via graph cuts? In *European Conference on Computer Vision*, pages 65–81, 2002.
- [57] N. Komodakis. Towards more efficient and effective LP-based algorithms for MRF optimization. In *European Conference on Computer Vision*, 2010.

REFERENCES

- [58] N. Komodakis and N. Paragios. Beyond loose lp-relaxations: Optimizing mrfs by repairing cycles. In *European Conference on Computer Vision*, 2008.
- [59] N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [60] N. Komodakis and N. Paragios. MRF-based blind image deconvolution. In *Asian Conference on Computer Vision*, 2012.
- [61] N. Komodakis, N. Paragios, and G. Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *International Conference on Computer Vision*, 2007.
- [62] N. Komodakis, N. Paragios, and G. Tziritas. Clustering via LP-based stabilities. In *Advances in Neural Information Processing Systems*, 2008.
- [63] N. Komodakis, N. Paragios, and G. Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2011.
- [64] N. Komodakis and G. Tziritas. A new framework for approximate labeling via graph cuts. In *International Conference on Computer Vision*, 2005.
- [65] N. Komodakis and G. Tziritas. Approximate labeling via graph-cuts based on linear programming. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007.
- [66] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Trans. on Image Processing*, 2007.
- [67] N. Komodakis and G. Tziritas. Morphable 3D-mosaics: a hybrid framework for photorealistic walkthroughs of large natural environments. *Advances in Multimedia*, 2008.
- [68] N. Komodakis and G. Tziritas. Real-time exploration and photorealistic reconstruction of large natural environments. *The Visual Computer*, 2008.
- [69] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic MRFs. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

REFERENCES

- [70] N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal dual strategies. *Computer Vision and Image Understanding*, 2008.
- [71] Nikos Komodakis. Efficient training for pairwise or higher order CRFs via dual decomposition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [72] Nikos Komodakis. Learning to cluster using high order graphical models with latent variables. In *International Conference on Computer Vision*, 2011.
- [73] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF optimization via dual decomposition: Message-passing revisited. In *International Conference on Computer Vision*, 2007.
- [74] Nikos Komodakis, Nikos Paragios, and Georgios Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010.
- [75] I. Kovtun. Partial optimal labeling search for a np-hard subclass of (max,+) problems. In *DAGM*, 2003.
- [76] M. P. Kumar, V. Kolmogorov, and P. H. S. Torr. An analysis of convex relaxations for MAP estimation. In *Advances in Neural Information Processing Systems*, 2007.
- [77] Sanjiv Kumar, Jonas August, and Martial Hebert. Exploiting inference for approximate parameter learning in discriminative fields: An empirical study. In *Intl. Conf. on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2005.
- [78] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 1996.
- [79] Xiangyang Lan, Stefan Roth, Daniel P. Huttenlocher, and Michael J. Black. Efficient belief propagation with learned higher-order markov random fields. In *European Conference on Computer Vision*, 2006.
- [80] T. Larsson, M. Patriksson, and A. Stromberg. Ergodic primal convergence in dual subgradient schemes for convex programming. *Mathematical Programming*, 86:283–312, 1999.

REFERENCES

- [81] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. A sparse texture representation using local affine regions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005.
- [82] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [83] A. Levin, Y. Weiss, F. Durand, and W.T. Freeman. Understanding and evaluating blind deconvolution algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1964–1971, 2009.
- [84] Anat Levin. Blind motion deblurring using image statistics. In *Advances in Neural Information Processing Systems*, 2006.
- [85] Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. Image and depth from a conventional camera with a coded aperture. In *Proc. ACM Siggraph*, 2007.
- [86] Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 3rd edition, 2009.
- [87] Yunpeng Li and Daniel P. Huttenlocher. Learning for stereo vision using the structured support vector machine. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [88] Talya Meltzer, Chen Yanover, and Yair Weiss. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In *International Conference on Computer Vision*, 2005.
- [89] Ofer Meshi, David Sontag, Tommi Jaakkola, and Amir Globerson. Learning efficiently with approximate inference via dual losses. In *Proc. Int. Conf. on Machine Learning*, 2010.
- [90] Daniel Munoz, J. Andrew (Drew) Bagnell, Nicolas Vandapel, and Martial Hebert. Contextual classification with functional max-margin markov networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [91] A. Nedic and D. P. Bertsekas. Incremental subgradient methods for non-differentiable optimization. *SIAM Journal on Optimization*, 12:109–138, 2001.

REFERENCES

- [92] Angelia Nedic and Asuman Ozdaglar. Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization*, 19:1757–1780, 2009.
- [93] R. Neji, A. Besbes, N. Komodakis, J-F. Deux, M. Maatou, A. Rahmouni, G. Bassez, G. Fleury, and N. Paragios. Clustering of the human skeletal muscle fibers using linear programming and angular hilbertian metrics. In *International Conference on Information Processing in Medical Imaging*, 2009.
- [94] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, 2001.
- [95] Timo Ojala, Topi Mäenpää, Matti Pietikäinen, Jaakko Viertola, Juha Kyllönen, and Sami Huovinen. Outex - new framework for empirical evaluation of texture analysis algorithms. In *International Conference on Pattern Recognition*, 2002.
- [96] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002.
- [97] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Computer Vision*, 2001.
- [98] Christos H. Papadimitriou and Kenneth Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Prentice-Hall, 1982.
- [99] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.
- [100] Brian Potetz. Efficient belief propagation for vision using linear constraint nodes. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [101] Ramesh Raskar, Amit Agrawal, and Jack Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. *Proc. ACM Siggraph*, pages 795–804, 2006.
- [102] Nathan Ratliff, J. Andrew (Drew) Bagnell, and Martin Zinkevich. (online) subgradient methods for structured prediction. In *International Conference on Artificial Intelligence and Statistics*, 2007.

REFERENCES

- [103] Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [104] Carsten Rother, Pushmeet Kohli, Wei Feng, and Jiaya Jia. Minimizing sparse higher order energy functions of discrete variables. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [105] Carsten Rother, Vladimir Kolmogorov, Victor Lempitsky, and Martin Szummer. Optimizing binary MRFs via extended roof duality. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [106] M. Sardaescu, N. Paragios, N. Komodakis, R. Raymond, P. Hernigou, and A. Rahmouni. Knee reconstruction through efficient linear programming. In *IEEE International Symposium on Biomedical Imaging*, 2008.
- [107] Daniel Scharstein and Chris Pal. Learning conditional random fields for stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [108] M.I. Schlesinger and V.V. Giginyak. Solution to structural recognition (MAX,+)-problems by their equivalent transformations. *Control Systems and Computers*, 2007.
- [109] Shai Shalev-Shwartz, Yoram Singer, and Nathan Srebro. Pegasos: Primal estimated sub-gradient solver for svm. In *Proc. Int. Conf. on Machine Learning*, 2007.
- [110] Qi Shan, Jiaya Jia, and Aseem Agarwala. High-quality motion deblurring from a single image. *Proc. ACM Siggraph*, 2008.
- [111] Alexander Shekhovtsov, Ivan Kovtun, and Václav Hlaváček. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 2008.
- [112] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [113] N.Z. Shor. *Minimization methods for nondifferentiable functions*. Springer, Berlin, 1985.
- [114] D. Sontag, T. Meltzer, A. Globerson, Y. Weiss, and T. Jaakkola. Tightening lp relaxations for map using message passing. In *Conference on Uncertainty in Artificial Intelligence*, 2008.

REFERENCES

- [115] David Sontag and Tommi Jaakkola. New outer bounds on the marginal polytope. In *Advances in Neural Information Processing Systems*, 2008.
- [116] A. Sotiras, N. Komodakis, B. Glocker, and N. Paragios. Graphical models and deformable diffeomorphic population registration using global and local metrics. In *Medical Image Computing and Computer-Assisted Intervention*, 2009.
- [117] A. Sotiras, N. Komodakis, G. Langs, and N. Paragios. Atlas-based deformable mutual population segmentation. In *IEEE International Symposium on Biomedical Imaging*, 2009.
- [118] A. Sotiras, R. Neji, J. Deux, N. Komodakis, G. Fleury, and N. Paragios. A kernel-based graphical model for diffusion tensor registration. In *IEEE International Symposium on Biomedical Imaging*, 2010.
- [119] Petter Strandmark, Fredrik Kahl, and Thomas Schoenemann. Parallel and distributed vision algorithms using dual decomposition. *Computer Vision and Image Understanding*, 2011.
- [120] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields. In *European Conference on Computer Vision*, 2006.
- [121] Martin Szummer, Pushmeet Kohli, and Derek Hoiem. Learning CRFs using graph cuts. In *European Conference on Computer Vision*, 2008.
- [122] Ben Taskar, Vassil Chatalbashev, and Daphne Koller. Learning associative markov networks. In *Proc. Int. Conf. on Machine Learning*, 2004.
- [123] Ben Taskar, Carlos Guestrin, and Daphne Koller. Max-margin markov networks. In *Advances in Neural Information Processing Systems*, 2004.
- [124] Manik Varma and Andrew Zisserman. A statistical approach to texture classification from single images. *Int. J. Computer Vision*, 2005.
- [125] V. Vazirani. *Approximation Algorithms*. Springer, 2001.
- [126] D. Verma and M. Meila. A comparison of spectral clustering algorithms. Technical report, 2001.
- [127] C. Wachinger, B. Glocker, J. Zeltner, N. Paragios, N. Komodakis, M.S. Hansen, and N. Navab. Deformable mosaicing for whole-body MRI. In *Medical Image Computing and Computer-Assisted Intervention*, 2008.

REFERENCES

- [128] M. Wainwright, T. Jaakkola, and A. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *IEEE Trans. on Information Theory*, 2005.
- [129] Y.L. Wang, J.F. Yang, W.T. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 2008.
- [130] Tomas Werner. A linear programming approach to max-sum problem: A review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007.
- [131] Tomas Werner. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (map-mrf). 2008.
- [132] Tomas Werner. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimization (MAP-MRF). In *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [133] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *European Conference on Computer Vision*, pages 157–170, 2010.
- [134] Chen Yanover, Talya Meltzer, and Yair Weiss. Linear Programming Relaxations and Belief Propagation – An Empirical Study. *J. Machine Learning Research*, 7, 2006.
- [135] J. Yedidia, W. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 2005.
- [136] Chun-Nam John Yu and Thorsten Joachims. Learning structural svms with latent variables. In *Proc. Int. Conf. on Machine Learning*, 2009.
- [137] D. Zikic, B. Glocker, O. Kutter, M. Groher, N. Komodakis, A. Kamen, N. Paragios, and N. Navab. Markov random field optimization for intensity-based 2D-3D registration. In *SPIE Medical Imaging*, 2010.
- [138] D. Zikic, B. Glocker, O. Kutter, M. Groher, N. Komodakis, A. Khamene, N. Paragios, and N. Navab. Linear intensity-based image registration by markov random fields and discrete optimization. *Medical Image Analysis*, 2010.